

A4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/46443 A2

(51) International Patent Classification⁷: C12N 15/57, 9/48, A01K 67/00, C07K 16/40, C12Q 1/68, A61K 38/48, G01N 33/50

(21) International Application Number: PCT/US00/34811

(22) International Filing Date:
19 December 2000 (19.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/172,055 23 December 1999 (23.12.1999) US
60/177,334 21 January 2000 (21.01.2000) US
60/178,884 28 January 2000 (28.01.2000) US
60/179,903 2 February 2000 (02.02.2000) US

#3, Sunnyvale, CA 94086 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). NGUYEN, Danniel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YANG, Jun-ming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue,

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/46443 A2

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

PROTEASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of 5 these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

10

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of 15 peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

20

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-dipeptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on 25 active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include 30 the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the active

catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) *Meth. Enzymol.* 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaeabacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, *supra*).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and prolylendopeptidase families

(Tan, F. et al. (1993) *J. Biol. Chem.* 268:16631-16638). The protease neutrophil may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) *J. Neurosci.* 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) *Neurology* 53:14-19) and myocardial infarction (Ross, A.M.

5 (1999) *Clin. Cardiol.* 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) *Aliment. Pharmacol. Ther.* 14:257-266; Rice, K.D.

10 et al. (1998) *Curr. Pharm. Des.* 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) *Urology* 33:11-16).

15 The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serve in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, 20 activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) *J. Biol. Chem.* 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

25 Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but 30 is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) *Adv. Neurol.* 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all

types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) *Cell* 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is

5 hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle

10 syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) *Annu. Rev. Med.* 50:57-74). A murine proto-oncogene, *Ubp*, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) *Oncogene* 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell

15 line to a non-dividing mature state (Maki, A. et al. (1996) *Differentiation* 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) *J. Pathol.* 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing

20 inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) *Curr. Opin. Chem. Biol.* 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see

Rawlings, N.D. and A.J. Barrett (1994) *Meth. Enzymol.* 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of 5 these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium

10 (Keyszer, G.M. (1995) *Arthritis Rheum.* 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit 15 also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) *J. Neurosci. Res.* 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic 20 neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, *supra*). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) *J. Neurotrauma* 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. 25 (1999) *J. Neurol. Sci.* 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic 30 mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention)

and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into 5 self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, *supra*; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. 10 Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, *supra*). Cowpox virus and 15 baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, *supra*; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

20 Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2-3, at which one of the aspartate residues is ionized and the other neutral. 25 The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer 30 contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum. Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease

states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) *Crit. Rev. Oncol.* 4:95-114).

5 **Metalloproteases**

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases 10 and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, 15 I. et al. (1998) *Horm. Metab. Res.* 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neuropeptides (Serizawa, A. et al. (1995) *J. Biol. Chem.* 270:2092-2098). Neuropeptides are a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) *Neuropeptides* 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade 20 components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) *Trends Neurosci.* 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexin-like domains can be used 25 to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2} -cysteine interaction, or "cysteine switch," exposing the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by 30 stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) *Trends Neurosci.* 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) *J. Clin. Invest.* 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) *Circulation* 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) *Am. J. Path.* 152:703), non-healing wounds

(Saarialho-Kere, U.K. et al. (1994) *J. Clin. Invest.* 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) *J. Cell Sci.* 108:3649), age-related macular degeneration (Steen, B. et al. (1998) *Invest. Ophthalmol. Vis. Sci.* 39:2194), emphysema (Finlay, G.A. et al. (1997) *Thorax* 52:502), myocardial infarction (Rohde, L.E. et al. (1999) *Circulation* 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) *Circulation* 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) *Cancer Res.* 56:2815; Anderson et al. (1996) *Cancer Res.* 56:715-718; Volpert, O.V. et al. (1996) *J. Clin. Invest.* 98:671; Taraboletti, G. et al. (1995) *J. NCI* 87:293; Davies, B. et al. (1993) *Cancer Res.* 53:2087). MMPs may be active in Alzheimer's disease.

10 A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, *supra*).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

20 ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) *J. Cell. Sci.* 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in *Drosophila* neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, *supra*). TACE has also been identified as the TNF activating enzyme (Black, R.A. et al. (1997) *Nature* 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound 25 pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

30 The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al.

(1997) *J. Biol. Chem.* 272:556). To date eleven members are recognized by the Human Genome Organization (HUGO; <http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved>). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) *Science* 284:1664; Abbaszade, I. (1999) *J. Biol. Chem.* 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., *supra*) and/or procollagen processing (Colige, A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2374).

10 **Protease inhibitors**

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) *Agents Actions Suppl.* 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al. (1995) *Biol. Biochem. Hoppe Seyler* 376:71-80). Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin, that binds the proenzyme, proacrosin, and thereby aides in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) *J. Biol. Chem.* 269:10133-10140). The Kunitz family of serine protease inhibitors are characterized by one or more "Kunitz domains" containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter- α -trypsin inhibitor, and bikunin. (Marlor, C.W. et al. (1997) *J. Biol. Chem.* 272:12202-12208.) Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin. Aprotinin has clinical utility in reduction of perioperative blood loss.

The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7," "PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," and "PRTS-14." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from

5 the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

10 14. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-14. In another alternative, the polynucleotide is selected

15 from the group consisting of SEQ ID NO:15-28.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism

20 comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c)

a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide 5 comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid 10 sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence 15 selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous 20 nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence 25 identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions 30 whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said

target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, c) a 5 polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

10 The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group 15 consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need 20 of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the 25 group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound 30 identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting 35 of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally

occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method

5 comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment

10 the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group

15 consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the

20 polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the

25 group consisting of SEQ ID NO:1-14, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test

30 compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in

35 altering expression of a target polynucleotide, wherein said target polynucleotide comprises a

sequence selected from the group consisting of SEQ ID NO:15-28, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said 5 method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to 10 a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific 15 hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a 20 polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of 25 a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

25

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank 30 homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each polypeptide.

35

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each

polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

5 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood 10 that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," 15 and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings 20 as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, 25 protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and 30 human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to 5 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions, 10 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The 15 encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and 20 arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, 25 polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. 30 Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by 35 directly interacting with PRTS or by acting on components of the biological pathway in which PRTS

participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments 5 containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

10 The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to 15 elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified 20 sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or 25 translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, 30 to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be 5 employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated 10 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to 15 produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least 20 interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
25	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
30	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
35	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

5 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one 10 biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

15 A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 20 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the 25 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:15-28 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:15-28, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:15-28 is useful, for 30 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:15-28 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:15-28 and the region of SEQ ID NO:15-28 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-14 is encoded by a fragment of SEQ ID NO:15-28. A fragment 35 of SEQ ID NO:1-14 comprises a region of unique amino acid sequence that specifically identifies

SEQ ID NO:1-14. For example, a fragment of SEQ ID NO:1-14 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-14. The precise length of a fragment of SEQ ID NO:1-14 and the region of SEQ ID NO:1-14 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended 5 purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two 10 or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore 15 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. 20 and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

25 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis 30 programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST 35 programs are commonly used with gap and other parameters set to default settings. For example, to

compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

5 *Matrix: BLOSUM62*
 Reward for match: 1
5 *Penalty for mismatch: -2*
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 Expect: 10
 Word Size: 11
10 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such 15 lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

25 The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment 30 program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

5 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

10 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

15 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

20 The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

25 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the 5 target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention 10 include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as 15 formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

20 The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells 25 or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

30 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of

PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

5 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

10 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

15 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

20 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

25 "Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

30 "Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

35 Probes and primers as used in the present invention typically comprise at least 15 contiguous

nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

5 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR 10 Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

15 Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South 20 West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for 25 microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences.

30 Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to

those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the 5 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

10 Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). 15 Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and 20 other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

25 The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a 30 protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will

reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which 5 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, 10 chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient 15 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" 20 includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to 25 animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in 30 vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), 35 supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 5 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may 10 possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in 15 which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) 20 set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

25 The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide 30 sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an 35 Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:15-28 or that distinguish between SEQ ID NO:15-28 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7032724H1 is the identification number of an Incyte cDNA sequence, and BRAXTDR12 is the cDNA library from which

it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70152356V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5364348) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6436155_002.edit is the identification number of a Genscan-predicted coding sequence, with g6436155 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon-stretching” algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

15 Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

20 The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

25 The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:15-28, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

30 The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28 which has at

least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:15-28. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in
10 accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its
15 derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of
20 RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using
25 reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:15-28 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and
30 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of

DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is

5 automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a

10 variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligation may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been

size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

5 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate
10 software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which
15 encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally
20 known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites,
25 alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULAR BREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or
30 improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of
35 DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"

breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

5 maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively,

10 PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid

15 sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

20 The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a

25 suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where

30 sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural

35 and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate

for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.*

264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors 5 containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) 10 Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be 15 used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

20 In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. 25 Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are 30 constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous

expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the 5 introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; 10 Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers 15 resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) 20 Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β 25 glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing 25 sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express 30 PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRTS using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humanà Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and 10 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to 15 synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, 20 cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing 25 polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, 30 lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing

of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) *Current Protocols in Immunology* 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E.*

coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the

resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with gastrointestinal, epithelial, reproductive, cardiovascular, cancerous, and inflamed tissues, and with normal kidney and normal skin tissues. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, 30 dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory- 35 Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel

syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis

5 hepatitis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon

10 angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart

15 disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-

20 ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,

25 osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative

30 disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract,

heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantar, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial

insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive 5 supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine 10 fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, 15 impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased 20 expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

25 In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such 30 disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J.*

Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate 5 antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be 10 generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, 15 G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy 20 identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such 25 immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques 30 may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The K_a determined for

a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

35 In another embodiment of the invention, polynucleotides encoding PRTS may be used for

somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

5 (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii)

10 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; 25 Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, 30 PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998)

Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous 5 gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method 10 (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the 15 polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. 20 Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. 25 et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the 30 return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to

the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are 5 described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver 10 polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has 15 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under 20 the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids 25 containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, 30 Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g.,

protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in 5 hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may 10 require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes 15 inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by 20 preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze 25 endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for 30 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for

chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize 5 complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be 10 extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds 15 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of 20 polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

25 At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound 30 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by 35 any method commonly known in the art. Typically, the expression of a specific nucleotide is detected

by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a

5 test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 10 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for

15 use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

20 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient.

25 Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes

30 including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case

35 of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting

formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

5 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

10 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et 15 al. (1999) *Science* 285:1569-1572).

20 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

25 A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is 30 preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active

moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending 5 on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their 10 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with 15 PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter 20 molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under 25 conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for 30 diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a 5 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject 10 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:15-28 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may 15 be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated 20 with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic 25 carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic 30 steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula,

atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

5 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an

10 autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus,

15 emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's

20 syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal

25 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder,

30 such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism,

hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, 5 dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, 10 chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, 15 trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, 20 bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental 25 retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic 30

disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of 5 prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, 10 impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

15 In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard 20 value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

25 In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with 30 values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

5 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further 10 progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be 15 employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are 20 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal 25 tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP 30 (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high

throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the 10 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify 15 genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the 20 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to 25 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of 30 transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies,

or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of 5 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a 10 signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested 15 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released 20 February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated 25 biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

30 Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given 35 conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and

analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are 5 visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein 10 spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

15 A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of 20 methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. 25 Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological 30 sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the

individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as

linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in

particular U.S. Ser. No. 60/172,055, U.S. Ser. No. 60/177,334, U.S. Ser. No. 60/178,884, and U.S. Ser. No. 60/179,903, are expressly incorporated by reference herein.

EXAMPLES

5 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. The Incyte cDNAs shown for SEQ ID NO:15 were derived from cDNA libraries constructed from small intestine, ovary, lung, skin, breast, prostate epithelium, and mixed myometrial tissues; umbilical cord blood, and teratocarcinoma cells which contained neuronal precursors. The Incyte cDNAs shown for SEQ ID NO:17 were derived from cDNA libraries constructed from a bronchial epithelium primary cell line, dermal microvascular endothelial cells, pancreas, ileum tissue associated with Crohn's disease, rib bone tissue associated with Patau's syndrome, kidney, thoracic dorsal root ganglion, and penis corpus cavernosum tissue. The Incyte cDNA shown for SEQ ID NO:18 was derived from a cDNA library constructed from brain tumor tissue. The Incyte cDNAs shown for SEQ ID NO:19 were derived from cDNA libraries constructed from adrenal gland, colon, and breast tissue. The Incyte cDNAs shown for SEQ ID NO:20 were derived from cDNA libraries constructed from T-lymphocytes, lung, breast, and penis corpus cavernosum tissues. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the

appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

5 PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

10 **II. Isolation of cDNA Clones**

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 15 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal 20 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

25 **III. Sequencing and Analysis**

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared 30 using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI 35 protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, 5 linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. 10 (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding 15 sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine 20 residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and 25 LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of 30 Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match

between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:15-28. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

30 "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory

and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity.

5 For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed
10 along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of
15 genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using
20 the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog,
25 the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

30 The sequences which were used to assemble SEQ ID NO:15-28 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:15-28 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available
35 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for

Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

15 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

20 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$25 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

30 The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The 35 product score represents a balance between fractional overlap and quality in a BLAST alignment. For

example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

5 Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; 10 embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, 15 cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

20 **VIII. Extension of PRTS Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using 25 OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

30 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme

(Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, 10 Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, 15 digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were 20 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides

designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:15-28 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is 5 specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine 10 triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon 15 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40 °C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

20 **X. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested 25 substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; 30 Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array

elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

10 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified 15 using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

20

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 30 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR 35 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and

coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

5 Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 10 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample 15 mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 20 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is 25 focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. 30 Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, 35 although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples 5 from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital 10 (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping 15 emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

20 XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 25 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

XII. Expression of PRTS

30 Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

35 Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic

resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS

5 by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA

10 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized

15 glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins

20 (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX, where applicable.

XIII. Functional Assays

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically

25 elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome

30 formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-

based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in 5 cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G.

10 (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using 15 magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PRTS Specific Antibodies

20 PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is 25 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-30 Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated 5 SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding 10 (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules 15 previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid 20 system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions 25 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. 30 Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are

performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

5 An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. Chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) *Anal. Biochem.* 247:305-309).

10 In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) *Gene* 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so 15 that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al: (1999) *FEBS Lett.* 447:53-57).

XVIII. Identification of PRTS Substrates

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal 25 extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then 30 amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) *J. Biol. Chem.* 272:16603-16609).

To screen for in vivo PRTS substrates, this method can be expanded to screen a cDNA

expression library displayed on the surface of phage particles (T7SELECT™10-3 Phage display vector, Novagen, Madison, WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad, CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PRTS Inhibitors

5 Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

10 In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) *Nat. Biotechnol.* 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for

15 PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

Although the invention has been described in connection with certain embodiments, it should be

20 understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Project ID	Polypeptide SEQ ID NO:	Polypeptide ID	Incye SEQ ID NO:	Polynucleotide SEQ ID NO:	Incye Polynucleotide ID
1714846	1	1714846CD1	15		1714846CB1
1856589	2	1856589CD1	16		1856589CB1
2617672	3	2617672CD1	17		2617672CB1
2769104	4	2769104CD1	18		2769104CB1
4802789	5	4802789CD1	19		4802789CB1
60116897	6	60116897CD1	20		60116897CB1
1866356	7	1866356CD1	21		1866356CB1
1872095	8	1872095CD1	22		1872095CB1
2278688	9	2278688CD1	23		2278688CB1
4043361	10	4043361CD1	24		4043361CB1
3937958	11	3937958CD1	25		3937958CB1
7257324	12	7257324CD1	26		7257324CB1
7472038	13	7472038CD1	27		7472038CB1
7472041	14	7472041CD1	28		7472041CB1

Table 2

Polyptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1714846	g6941890	0.0	Ubiquitin-specific processing protease [Mus musculus] (Valero, R. et al. (1999) Genomics 62:395-405)
2	1856589	g1143194	1.2e-45	Prostasin [Homo sapiens] (Yu, J.X. et al. (1994) J. Biol. Chem. 269:18843-18848)
3	2617672	g4929827	8.0e-118	Tubulo-interstitial nephritis antigen TIN-Ag [Mus musculus]
4	2769104	g179644	3.3e-28	Human complement C1r [Homo sapiens]
5	4802789	g4454565	4.1e-30	Ubiquitin Processing protease [Homo sapiens] (Cai, S. et al. Proc. Natl. Acad. Sci. USA (1999) 96:2828-2833)
6	60116897	g9886747	0.0	VEGF induced aminopeptidase [Mus musculus]
7	1866356CD1	g20888823	1.5e-68	Similarity to the peptidase family A2 [Caenorhabditis elegans]
8	1872095CD1	g2347100	1.7e-22	Ubiquitin-specific protease [Arabidopsis thaliana]
9	2278688CD1	g1184161	0.0	Aminopeptidase [Mus musculus]
10	4043361CD1	g9843781	2.6e-104	Putative pyroglutamyl-peptidase I [Mus musculus]
11	3937958CD1	g180950	5.7e-16	Carboxylesterase [Homo sapiens]
12	7257324CD1	g2116650	1.1e-78	Alpha-1-antitrypsin [Cercopithecus aethiops] (Colau, B. et al. (1984) DNA 3:327-330; Yoshida, K. et al. (1999) J. BioChem. Mol. Biol. Biophys. 3:59-63)
13	7472038CD1	g293230	4.0e-106	Aspartic protease [Aedes aegypti]
14	7472041CD1	g3088553	1.1e-14	Cystatin-related epididymal spermatoenic protein [Homo sapiens] (Cornwall, G.A., Hsia, N., and Sutton, H.G. Biochem. J. (1999) 340 (pt 1):85-93)

Table 3

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1714846CD1	1055	S85 T95 S109 T123 S136 T147 T286 S357 S375 S467 S489 T541 S546 T557 S631 S632 S745 T796 T824 S835 S892 S945 S1021 S1032 T1050 T55 S113 T235 T267 S354 S460 S513 S582 S719 Y575 Y872 Y873	N31 N256 N561 N646 N833	Ubiquitin C-terminal hydrolase: L232-W703, K823-F899 Ubiquitin C-term. hydrolase signature 1: V169-Y200 Ubiquitin C-term. hydrolase signature 2: G170-L187, S258-T267, E617-R638, I587-N656, N173-N408, D562-G601	HMmer-PFAM BLIMPS- BLOCKS- BLAST- PRODOM BLAST-DOMO MOTIFS
2	1856589CD1	358	S47 T188 T5 S105 T143 Y247	N150	Chymotrypsin family: G115-C130, F173-V187, E277-A289 Trypsin family: W100-I1327, C114-C130, P314-I327, W100-M331, Trypsin family His active site: V125-C130, L106-N150 Trypsin family Ser active site: V265-K310 Kringle domain: C114-S131, I196-S217, G286-I327 Apple domain: G116-P148, V187-Q221, I270-W304, E305-N333	HMmer-PFAM BLIMPS- BLOCKS- BLIMPS- PRINTS ProfileScan BLAST- PRODOM BLAST-DOMO MOTIFS
3	2617672CD1	467	T80 T117 T126 T169 T205 S296 T411 T180 S210 S239 S401 T417	N78 N161	Signal peptide: M1-A19, M1-G21 Papain family protease: D222-W456, Q223-F232, T399-G408, Y420-H436, H400-E410, Y420-S426, F76-G457, D145-V455 Cys protease His active site: G398-G408 Tubulointerstitial nephritis antigen: G45-I193	HMmer SPScan HMmer-PFAM BLIMPS- BLOCKS- BLIMPS- PRINTS BLAST- PRODOM BLAST-DOMO MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	2769104CD1	187	S67 T162 S131 S134 T138	N147	CUB domain (extracellular domain found in complement proteins): G40-Y160 Complement C1r/C1s repeat: C36-V163, Q51-Y160, M24-Y160 Signal peptide: M1-A35	HHMER -PFAM BLAST -DOMO MOTIFS SPScan HHMER
5	4802789CD1	289	T118 S28 S109 T213 S236 S261 S117 S102 S108 S188 S225 T265 S271	N119 N186	Transmembrane domain: W25-L52 Ubiquitin C-term. hydrolase signature 1: G191-L208 Signal peptide: M1-A44	HHMER BLIMPS- BLOCKS HHMER -PFAM MOTIFS SPScan
6	60116897CD1	960	S225 S483 T57 T87 S124 T197 S321 T343 S357 T407 S502 S607 S701 S738 S744 S817 S906 S926 T933 S10 S94 T183 S221 T256 S303 S359 S432 S486 S558 S740 S781 T830 T951 Y312 Y622 Y679 Y885	N85 N103 N119 N219 N294 N405 N431 N650 N714 N879	Zn metallopeptidase family M1: L69-G458 Zn membrane alanyl dipeptidase: R205-F220, F253-Y268, F331-L341, V367-T382, W386-Y398 Neutral Zn-protease: W64-S500, G529-L837, T521-S899, W64-T902, P54-D555, K553-L837, V849-L956 Neutral Zn-protease, Zn binding region: V367-W376, V367-F377 Signal peptide: M1-C35	HHMER -PFAM BLIMPS- BLOCKS BLIMPS- PRINTS BLAST- PRODOM BLAST -DOMO MOTIFS SPScan
7	1866356CD1	525	S82 S90 T159 T174 S288 S290 T311 T356 S397 T479 S522 S107 S122 S165 S228		Signal peptide: M1-S26 Similarity to the peptidase family A2 PD138963: F157-G422	SPScan SPScan BLAST- PRODOM

Table 3 (cont.)

SEQ NO:	Incite ID	Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	1872095CD1	795	S274 Y523 S274 Y523 T693 T251 N443 N448 N536 N617 N670 N436 N711 N712 Y593-H610	S522 T251 S314 S332 T337 S377 S383 S392 S470 T472 S555 S557 N720 N788	N171 N381 N443 N448 N536 N617 N670 N436 N711 N712 Y593-H610	Ubiquitin carboxyl-terminal hydrolase 1 motif: G199-L213	MOTIFS

						Ubiquitin carboxyl-terminal hydrolase 2 motif: T198-L229	MOTIFS
						Ubiquitin carboxyl-terminal hydrolases family UCH-1: T198-L229	HMMER - PFAM
						Ubiquitin carboxyl-terminal hydrolases family UCH-2: K589-K701	HMMER - PFAM
						Protease ubiquitin hydrolase, ubiquitin-specific enzyme, deubiquitinating carboxyl-terminal thiolesterase, processing, conjugation: PD017412: S470-L541	BLAST- PRODOM

						Ubiquitin carboxyl-terminal hydrolases family 2: DM00659 P40818 782-1103: L203-D386	BLAST- DOMO
--	--	--	--	--	--	---	-------------

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	2278688CD1	919	T177 T325 Y326 S379 S427 S547 T548 S549 S632 T633 S667 T669 T721 T758 T759 S32 S33 T143 T325 Y326 S341 S342 S486 S522	N62 N484 N648	Membrane alanyl dipeptidase: PR00756: R185-F200, F235-V250, F313-L323, V349-T364, W368-W380, Zinc, aminopeptidase, metallopeptidase, neutral: DM00700 P164606 80-887: R53-Y842 Zinc protease: V349-Q357 Leucine zipper: I3-P23 Signal peptide: M1-S39	BLIMPS-PRINTS BLAST-DOMO HMMER
10	4043361CD1	209	S118	N22	Peptidase family M1: L54-G441 Aminopeptidase, hydrolase, metalloprotease, zinc, N-glycoprotein, transmembrane, signal, anchor, membrane: PD001134: R53-S486 Zinc, aminopeptidase, metallopeptidase, neutral: DM00700 P37898 1-794: E52-G845 Pyroglutamyl peptidase: K6-L182 Pyrrolidone carboxyl peptidase: PR00140: T11-L31, S66-E85 (P<0.0041)	HMMER-PFAM BLAST-PRODOM BLAST-DOMO HMMER-PFAM
11	3937958CD1	77	Y35 T47 S68		Peptidase, carboxylate, pyrrolidone, pyroglutamyl: DM03107 P42673 1-212: K6-G145	BLAST-DOMO
					Carboxylesterase domain: E4-W62 Esterase, hydrolase, precursor, signal, glycoprotein, serine, carboxylesterase family: PD00169: K3-W62	HMMER-PFAM BLAST-PRODOM
					Cholinesterase: DM00390 Q04791 355-538: K3-W62 Type B carboxylesterase: W15-N25	BLAST-DOMO BLIMPS-BLOCKS

Table 3 (cont.)

SEQ NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7257324CD1	414	S93 T94 T223 T258 T16 S26 T124 S182 S235 S300 S346 S396 Y118	N221 N233 N267	Serpins protein signatures BL00284: N71-T94, A173-T1193, T200-M241, V306-F332, D387-P411.	BLIMPS-BLOCKS
					Serpins signature: G364-K414	ProfileScan
					Serpin, serine protease inhibitor, signal, precursor, glycoprotein, plasma, proteinase: PD000192: A44-P411	BLAST-PRODOM
					Serpins: DM00112 P01009 47-413: D54-N410	BLAST-DOMO
					Signal peptide: M1-G19	HMmer SPSCAN
					Serpins (serine protease inhibitors): A45-P411	HMmer-PFAM
13	7472038CD1	397	S127 T166 S317 T381 S337 Y340 S16 S31 T90 T154 S252	N156 N166 N169 N178 N190 N195 N245 N298	Pepsin (A1) aspartic protease signature PR00792A: I84-V104, G230-T243, V278-L289, W369-D384	BLIMPS-PRINTS
					Aspartyl protease, hydrolase precursor, signal, zymogen, glycoprotein, multigene: P69-S307	BLAST-PRODOM
					Eukaryotic and viral aspartyl proteases: DM00126 Q03168 19-385: R23-A395	BLAST-DOMO
					Aspartyl protease: V93-V104, V278-L289	MOTIFS
					Eukaryotic aspartyl protease: P69-A395	HMmer-PFAM
					Eukaryotic and viral aspartic proteases BL00141: F91-S106, D184-S195, G235-G244, V278-L287, I370-A393	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7472041CD1	145	T76 S13 S19 S37 T83 S105	N42 N54 N57 N94 N98 N131 N132	Cysteine proteases, inhibitors: DM00182 P01035 1-110: G30-C134 Cysteine proteases inhibitor: R66-T89	BLAST-DOGO BLIMPS-BLOCKS HMMER SPScan

Table 4

Polynucleotide SEQ ID NO:	Incute Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
15	1714846CB1	4028	1349-1416, 1-199, 1903-3217	6831476H1 (SINTNOR01) 6773219H1 (OVARDIRO1) 6426758H1 (LUNGNON07) 1870084F6 (SKINB1T01) 898127H1 (BRSTN0T05) 643334H1 (LUNGNON07) 4442573H1 (SINTNOT22) 6286315H2 (EP1PUNA01) 1714846F6 (UCMCNOT02) 257076T6 (HNT2RAT01) 6487217H1 (MIXDNTNB01) 95364348	1 650 998 1575 1964 1999 2572 2586 3058 3300 3635 385	499 1271 1685 1995 2210 2596 2868 3110 3631 3988 4028 839
16	1856589CB1	1422	539-570, 324-395, 1-214, 756-933	70152356V1 70161001V1 70157441V1 60106256B2	1 359 686 976	569 824 1218 1422
17	2617672CB1	1911	1-619	548654H1 (BEPINT01) 2170381F6 (ENDCN0T03) 1437060F1 (PANCNOT08) 70098221V1 1428845H1 (SINTB5T01) 3290066H1 (BONRFFET01) 2994130H1 (KIDNFFET02) 3601537H1 (DRGTMNOT01) 3702672H1 (PENCN0T07) 754098R1 (BRAITUT02) 70186361V1	1 1 150 476 774 1170 1318 1423 1644 1760 1 143	675 268 1031 1352 1408 1569 1715 1850 1911 386 847
18	2769104CB1	854	1-176	70186120V1 3494839F6 (ADRETTUT07) 70005795D1 2630625T6 (COLNTUT15) 605612H1 (BRSTTUT01)	432 1 660 708 1198	854 685 1266 1364 1385
19	4802789CB1	1386	1-23, 343-503			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
20	60116897CB1	3323	2502-2610, 1122-1879	3154611F6 (TLYMTXT02) 60116918U1	1 740	834 1236
				2832568F6 (TLYMMNOT03) 2830930F7 (TLYMMNOT03) 6510679H1 (LUNGTU01)	1119 1610 1877	1657 2078 2180
				2849992F6 (BRSTTUT13) 3200003F6 (PENCNOT02)	2135	2641
				2849992T6 (BRSTTUT13) 3201617F6 (PENCNOT02)	2368 2792	2862 3323
21	1866356	2123	1-1590	824817R1 (PROSNOT06) 3257810H1 (PROSTUS08)	1 2004	551 2123
				5726464H1 (UTRSTUT05) 3739625T6 (MENTNOT01)	244 1669	904 2075
				258590R6 (HNT2RAT01) 6157882H1 (MONOTXN05)	642	1073
				6269726H1 (BRAFFEN03)	1821	2092
				4570803H1 (GBLADIT02)	1 1046	2092 21705
				144288LT6 (THYRNNOT01)	555	927
				1388162H1 (CARGDIT02)	2234	2893
				4662176H2 (BRSTTUT20)	1368 1545	1619 1809
				1344669H1 (PROSNOT11)	1714	1962
				SXBC01873V1	1898	2461
				449756R6 (TLYMMNOT02)	219	734
				449756T6 (TLYMMNOT02)	797	1459
				SXBC00314V1	1950	2515

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
23	2278688	4170	1-245, 3069-3624, 1149-1809	2254713H1 (OVARTUT01) 097483R1 (PITUNDR01) 3271744H1 (BRAINNOT20) 4422961H1 (BRAPIIT01) 1378162H1 (LUNGNOT10) 3076825H1 (BONEUNTO1) 3556490H1 (LUNGNOT31) 1368447H1 (SCORNON02) 4662177H2 (BRSTTUT20) 38533790H1 (BRAITUT12) 1877059H1 (LEUKNOT03) 1349282T1 (LATRITUT02) 4289627F6 (BRABDIRO1) 1289505T1 (BRAINNOT11) 2373989F6 (LISLNOT01) 097483R1 (PITUNDR01) 2698679H1 (UTRSNOT12) 2110561H1 (BRAITUT03) 3011419H1 (MUSCNOT07) 1394210H1 (THYRNOT03) 4880281H1 (UTRMTM01) 4043361F6 (LUNGNOT35) 6777288T1 (OVARDIRO1) 6121924H1 (BRAHNON05) 7032724H1 (BRAXTDR12) 4692968T6 (BRAENOT02) 18713340F6 (SKINB01) 3429631T6 (SKINNOT04) 7257324H1 (SKIRTDC01) GNN.96436155 002.edit GNN.95830433_004.edit	453 2611 1433 1258 1110 2127 759 3998 1 2411 2062 3814 72 3530 1570 2933 1817 669 2369 1001 524 1 436 1022 1 636 1256 416 1 1 1	710 3220 1667 1500 1309 2391 1064 4170 271 2705 2330 4157 578 4149 2094 3639 2134 950 2623 1296 767 593 1216 1538 480 1265 1497 1476 474 1194 438 438
24	4043361	767	1-66			
25	3937958	1538	385-506, 1-78, 1293-1538			
26	7257324	1497	651-770, 67-206			
27	7472038	1194	1-29, 788-1194			
28	7472041	438	1-27			

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
15	1714846CB1	LUNGNCN07
16	1856589CB1	PROSNCT18
17	2617672CB1	PANCNCT08
18	2769104CB1	COLANCT02
19	4802789CB1	ADRETUT07
20	60116897CB1	TLYMNCNCT03
21	1866356CB1	HNT2RAT01
22	1872095CB1	THYRNCT03
23	2278688CB1	LATRTUT02
24	4043361CB1	LUNGNCNCT35
25	3937958CB1	KIDNNCT05
26	7257324CB1	SKINNCT04

Table 6

Library	Vector	Library Description
LUNGNON07	PINCY	This normalized lung tissue library was constructed from RNA isolated from a lung tissue library. The library was normalized in two rounds using conditions adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232 and Bonaldo et al. (1996) Genome Res. 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSNOT18	PINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated adenofibromatous hyperplasia; this tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
PANCNOT08	PINCY	Library was constructed using RNA isolated from pancreatic tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology for the associated tumor tissue indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardivascular disease, benign neoplasm in the large bowel, and a cataract. Previous surgeries included a total splenectomy, cholecystectomy, and abdominal hysterectomy. Family history included cardivascular disease, type II diabetes, and stomach cancer.
COLANOT02	PINCY	Library was constructed using RNA isolated from diseased ascending colon tissue removed from a 25-year-old Caucasian female during a multiple segmental resection of the large bowel. Pathology indicated moderately to severely active chronic ulcerative colitis, involving the entire colectomy specimen and sparing 2 cm of the attached ileum. Grossly, the specimen showed continuous involvement from the rectum proximally; marked mucosal atrophy and no skip areas were identified. Microscopically, the specimen showed dense, predominantly mucosal inflammation and crypt abscesses. Patient history included benign large bowel neoplasm. Previous surgeries included a polypectomy.
ADRETUT07	PINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 43-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated pheochromocytoma.

Table 6 (cont.)

Library	Vector	Library Description
TLYMN03	PINCY	Library was constructed using RNA isolated from nonactivated Th1 cells. These cells were differentiated from umbilical cord CD4 T cells with IL-12 and B7-transfected COS cells.
HNT2RAT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
LATRTU02	PINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
LUNGNOT35	PINCY	Library was constructed using RNA isolated from lung tissue removed from a 62-year-old Caucasian female. Pathology for the associated tumor tissue indicated a grade 1 spindle cell carcinoma forming a nodule. Patient history included depression, thrombophlebitis, and hyperlipidemia. Family history included cerebrovascular disease, atherosclerotic coronary artery disease, breast cancer, colon cancer, type II diabetes, and malignant skin melanoma.
THYRNOT03	PINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
KIDNN05	PSPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
SKINNOT04	PINCY	Library was constructed using RNA isolated from breast skin tissue removed from a 70-year-old Caucasian female during a breast biopsy and resection.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3589-3402.	<i>EST</i> : Probability value= 1.0E-8 or less <i>Assembled EST</i> : fasta Identity= 95% or greater and <i>Match length</i> =200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and search.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>EST</i> : fasta E value=1.0E-6 <i>Assembled EST</i> : fasta Identity= 95% or greater and <i>Match length</i> =200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) <i>Our World View</i> , in a Nusshill, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-499; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
98	Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Person, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Person, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-14.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:15-28.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
 b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
 c) a polynucleotide sequence complementary to a),
 d) a polynucleotide sequence complementary to b), and
10 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
20 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

 a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

18. A method for treating a disease or condition associated with decreased expression of 5 functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15 21. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 20.

20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 23.

30 25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in 10 the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

- 15 comprising:
 - a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
 - b) detecting altered expression of the target polynucleotide, and
 - c) comparing the expression of the target polynucleotide in the presence of varying amounts of 20 the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at 25 least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- 30 d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.
YANG, Junming
BAUGHN, Mariah R.
BURFORD, Neil
AU-YOUNG, Janice
LU, Dyung Aina M.
REDDY, Roopa
YUE, Henry
NGUYEN, Danniel B.
TANG, Y. Tom
YAO, Monique G.
LAL, Preeti

<120> PROTEASES

<130> PI-0003 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/172,055; 60/177,334; 60/178,884; 60/179,903
<151> 1999-12-23; 2000-01-21; 2000-01-28; 2000-02-02

<160> 28

<170> PERL Program

<210> 1
<211> 1055
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 1714846CD1

<400> 1

Met Thr Val Glu Gln Asn Val Leu Gln Gln Ser Ala Ala Gln Lys
1 5 10 15
His Gln Gln Thr Phe Leu Asn Gln Leu Arg Glu Ile Thr Gly Ile
20 25 30
Asn Asp Thr Gln Ile Leu Gln Gln Ala Leu Lys Asp Ser Asn Gly
35 40 45
Asn Leu Glu Leu Ala Val Ala Phe Leu Thr Ala Lys Asn Ala Lys
50 55 60
Thr Pro Gln Gln Glu Glu Thr Thr Tyr Tyr Gln Thr Ala Leu Pro
65 70 75
Gly Asn Asp Arg Tyr Ile Ser Val Gly Ser Gln Ala Asp Thr Asn
80 85 90
Val Ile Asp Leu Thr Gly Asp Asp Lys Asp Asp Leu Gln Arg Ala
95 100 105
Ile Ala Leu Ser Leu Ala Glu Ser Asn Arg Ala Phe Arg Glu Thr
110 115 120
Gly Ile Thr Asp Glu Glu Gln Ala Ile Ser Arg Val Leu Glu Ala
125 130 135
Ser Ile Ala Glu Asn Lys Ala Cys Leu Lys Arg Thr Pro Thr Glu
140 145 150
Val Trp Arg Asp Ser Arg Asn Pro Tyr Asp Arg Lys Arg Gln Asp
155 160 165
Lys Ala Pro Val Gly Leu Lys Asn Val Gly Asn Thr Cys Trp Phe
170 175 180
Ser Ala Val Ile Gln Ser Leu Phe Asn Leu Leu Glu Phe Arg Arg
185 190 195
Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala Gln Asp Leu Pro
200 205 210
Arg Asn Gln Lys Glu His Arg Asn Leu Pro Phe Met Arg Glu Leu
215 220 225
Arg Tyr Leu Phe Ala Leu Leu Val Gly Thr Lys Arg Lys Tyr Val

	230	235	240
Asp Pro Ser Arg	Ala Val Glu Ile Leu Lys	Asp Ala Phe Lys	Ser
245	250	255	
Asn Asp Ser Gln	Gln Gln Asp Val Ser	Glu Phe Thr His Lys	Leu
260	265	270	
Leu Asp Trp Leu	Glu Asp Ala Phe Gln	Met Lys Ala Glu Glu	Glu
275	280	285	
Thr Asp Glu Glu	Lys Pro Lys Asn Pro	Met Val Glu Leu Phe	Tyr
290	295	300	
Gly Arg Phe Leu	Ala Val Gly Val Leu	Glu Gly Lys Lys Phe	Glu
305	310	315	
Asn Thr Glu Met	Phe Gly Gln Tyr Pro	Leu Gln Val Asn Gly	Phe
320	325	330	
Lys Asp Leu His	Glu Cys Leu Glu Ala	Ala Met Ile Glu Gly	Glu
335	340	345	
Ile Glu Ser Leu	His Ser Glu Asn Ser	Gly Lys Ser Gly Gln	Glu
350	355	360	
His Trp Phe Thr	Glu Leu Pro Pro Val	Leu Thr Phe Glu Leu	Ser
365	370	375	
Arg Phe Glu Phe	Asn Gln Ala Leu Gly	Arg Pro Glu Lys Ile	His
380	385	390	
Asn Lys Leu Glu	Phe Pro Gln Val Leu	Tyr Leu Asp Arg Tyr	Met
395	400	405	
His Arg Asn Arg	Glu Ile Thr Arg Ile	Lys Arg Glu Glu Ile	Lys
410	415	420	
Arg Leu Lys Asp	Tyr Leu Thr Val Leu	Gln Gln Arg Leu Glu	Arg
425	430	435	
Tyr Leu Ser Tyr	Gly Ser Gly Pro Lys	Arg Phe Pro Leu Val	Asp
440	445	450	
Val Leu Gln Tyr	Ala Leu Glu Phe Ala	Ser Ser Lys Pro Val	Cys
455	460	465	
Thr Ser Pro Val	Asp Asp Ile Asp Ala	Ser Ser Pro Pro Ser	Gly
470	475	480	
Ser Ile Pro Ser	Gln Thr Leu Pro Ser	Thr Thr Glu Gln Gln	Gly
485	490	495	
Ala Leu Ser Ser	Glu Leu Pro Ser Thr	Ser Pro Ser Ser Val	Ala
500	505	510	
Ala Ile Ser Ser	Arg Ser Val Ile His	Lys Pro Phe Thr Gln	Ser
515	520	525	
Arg Ile Pro Pro	Asp Leu Pro Met His	Pro Ala Pro Arg His	Ile
530	535	540	
Thr Glu Glu Glu	Leu Ser Val Leu Glu	Ser Cys Leu His Arg	Trp
545	550	555	
Arg Thr Glu Ile	Glu Asn Asp Thr Arg	Asp Leu Gln Glu Ser	Ile
560	565	570	
Ser Arg Ile His	Arg Thr Ile Glu Leu	Met Tyr Ser Asp Lys	Ser
575	580	585	
Met Ile Gln Val	Pro Tyr Arg Leu His	Ala Val Leu Val His	Glu
590	595	600	
Gly Gln Ala Asn	Ala Gly His Tyr Trp	Ala Tyr Ile Phe Asp	His
605	610	615	
Arg Glu Ser Arg	Trp Met Lys Tyr Asn	Asp Ile Ala Val Thr	Lys
620	625	630	
Ser Ser Trp Glu	Glu Leu Val Arg Asp	Ser Phe Gly Gly Tyr	Arg
635	640	645	
Asn Ala Ser Ala	Tyr Cys Leu Met Tyr	Ile Asn Asp Lys Ala	Gln
650	655	660	
Phe Leu Ile Gln	Glu Glu Phe Asn Lys	Glu Thr Gly Gln Pro	Leu
665	670	675	
Val Gly Ile Glu	Thr Leu Pro Pro Asp	Leu Arg Asp Phe Val	Glu
680	685	690	
Glu Asp Asn Gln	Arg Phe Glu Lys Glu	Leu Glu Glu Trp Asp	Ala
695	700	705	
Gln Leu Ala Gln	Lys Ala Leu Gln Glu	Lys Leu Leu Ala Ser	Gln
710	715	720	
Lys Leu Arg Glu	Ser Glu Thr Ser Val	Thr Ala Gln Ala	Ala
725	730	735	

Gly Asp Pro Glu Tyr Leu Glu Gln Pro Ser Arg Ser Asp Phe Ser
 740 745 750
 Lys His Leu Lys Glu Glu Thr Ile Gln Ile Ile Thr Lys Ala Ser
 755 760 765
 His Glu His Glu Asp Lys Ser Pro Glu Thr Val Leu Gln Ser Ala
 770 775 780
 Ile Lys Leu Glu Tyr Ala Arg Leu Val Lys Leu Ala Gln Glu Asp
 785 790 795
 Thr Pro Pro Glu Thr Asp Tyr Arg Leu His His Val Val Val Tyr
 800 805 810
 Phe Ile Gln Asn Gln Ala Pro Lys Lys Ile Ile Glu Lys Thr Leu
 815 820 825
 Leu Glu Gln Phe Gly Asp Arg Asn Leu Ser Phe Asp Glu Arg Cys
 830 835 840
 His Asn Ile Met Lys Val Ala Gln Ala Lys Leu Glu Met Ile Lys
 845 850 855
 Pro Glu Glu Val Asn Leu Glu Glu Tyr Glu Glu Trp His Gln Asp
 860 865 870
 Tyr Arg Lys Phe Arg Glu Thr Thr Met Tyr Leu Ile Ile Gly Leu
 875 880 885
 Glu Asn Phe Gln Arg Glu Ser Tyr Ile Asp Ser Leu Leu Phe Leu
 890 895 900
 Ile Cys Ala Tyr Gln Asn Asn Lys Glu Leu Leu Ser Lys Gly Leu
 905 910 915
 Tyr Arg Gly His Asp Glu Glu Leu Ile Ser His Tyr Arg Arg Glu
 920 925 930
 Cys Leu Leu Lys Leu Asn Glu Gln Ala Ala Glu Leu Phe Glu Ser
 935 940 945
 Gly Glu Asp Arg Glu Val Asn Asn Gly Leu Ile Ile Met Asn Glu
 950 955 960
 Phe Ile Val Pro Phe Leu Pro Leu Leu Leu Val Asp Glu Met Glu
 965 970 975
 Glu Lys Asp Ile Leu Ala Val Glu Asp Met Arg Asn Arg Trp Cys
 980 985 990
 Ser Tyr Leu Gly Gln Glu Met Glu Pro His Leu Gln Glu Lys Leu
 995 1000 1005
 Thr Asp Phe Leu Pro Lys Leu Leu Asp Cys Ser Met Glu Ile Lys
 1010 1015 1020
 Ser Phe His Glu Pro Pro Lys Leu Pro Ser Tyr Ser Thr His Glu
 1025 1030 1035
 Leu Cys Glu Arg Phe Ala Arg Ile Met Leu Ser Leu Ser Arg Thr
 1040 1045 1050
 Pro Ala Asp Gly Arg
 1055

<210> 2
 <211> 358
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1856589CD1

<400> 2
 Met Gly Ala Ala Thr Cys Arg Gly Ser Arg Ile Pro Ser Gly Pro
 1 5 10 15
 Pro Val Gln Gly Glu Arg Ser Ala Pro Arg Phe Gly Val Thr Ser
 20 25 30
 Leu Ser Leu Trp Pro Ala Asp Phe Lys Asp Asn Trp Arg Ile Ala
 35 40 45
 Gly Ser Arg Gln Glu Val Ala Leu Ala Gly Glu Pro Ala Asp Gln
 50 55 60
 Gln Gln Thr His Leu Arg Arg Leu Pro Tyr Arg Gln Thr Leu Gly
 65 70 75
 Tyr Lys Glu Asp Thr Thr Asn Pro Val Cys Gly Glu Pro Trp Trp
 80 85 90

Ser Glu Asp Leu Glu Met Thr Arg His Trp Pro Trp Glu Val Ser
 95 100 105
 Leu Arg Met Glu Asn Glu His Val Cys Gly Gly Ala Leu Ile Asp
 110 115 120
 Pro Ser Trp Val Val Thr Ala Ala His Cys Ser Gln Gly Thr Lys
 125 130 135
 Glu Tyr Ser Val Val Leu Gly Thr Ser Lys Leu Gln Pro Met Asn
 140 145 150
 Phe Ser Arg Ala Leu Trp Val Pro Val Arg Asp Ile Ile Met His
 155 160 165
 Pro Lys Tyr Trp Gly Arg Ala Phe Ile Met Gly Asp Val Ala Leu
 170 175 180
 Val His Leu Gln Thr Pro Val Thr Phe Ser Glu Tyr Val Gln Pro
 185 190 195
 Ile Cys Leu Pro Glu Pro Asn Phe Asn Leu Lys Val Gly Thr Gln
 200 205 210
 Cys Trp Val Thr Gly Trp Ser Gln Val Lys Gln Arg Phe Ser Gly
 215 220 225
 Ser Thr Ala Asn Ser Met Leu Thr Pro Glu Leu Gln Glu Ala Glu
 230 235 240
 Val Phe Ile Met Asp Asn Lys Arg Cys Asp Arg His Tyr Lys Lys
 245 250 255
 Ser Phe Phe Pro Leu Val Val Pro Leu Val Leu Gly Asp Met Ile
 260 265 270
 Cys Ala Thr Asn Tyr Gly Glu Asn Leu Cys Tyr Gly Asp Ser Gly
 275 280 285
 Gly Pro Leu Ala Cys Glu Val Glu Gly Arg Trp Ile Leu Ala Gly
 290 295 300
 Val Leu Ser Trp Glu Lys Ala Cys Val Lys Ala Gln Asn Pro Gly
 305 310 315
 Val Tyr Thr Arg Val Thr Lys Tyr Thr Lys Trp Ile Lys Lys Gln
 320 325 330
 Met Ser Asn Gly Ala Phe Ser Gly Pro Cys Ala Ser Ala Cys Leu
 335 340 345
 Leu Phe Leu Cys Trp Pro Leu Gln Pro Gln Met Gly Ser
 350 355

<210> 3
 <211> 467
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2617672CD1

<400> 3
 Met Trp Arg Cys Pro Leu Gly Leu Leu Leu Leu Pro Leu Ala
 1 5 10 15
 Gly His Leu Ala Leu Gly Ala Gln Gln Gly Arg Gly Arg Arg Glu
 20 25 30
 Leu Ala Pro Gly Leu His Leu Arg Gly Ile Arg Asp Ala Gly Gly
 35 40 45
 Arg Tyr Cys Gln Glu Gln Asp Leu Cys Cys Arg Gly Arg Ala Asp
 50 55 60
 Asp Cys Ala Leu Pro Tyr Leu Gly Ala Ile Cys Tyr Cys Asp Leu
 65 70 75
 Phe Cys Asn Arg Thr Val Ser Asp Cys Cys Pro Asp Phe Trp Asp
 80 85 90
 Phe Cys Leu Gly Val Pro Pro Pro Phe Pro Pro Ile Gln Gly Cys
 95 100 105
 Met His Gly Gly Arg Ile Tyr Pro Val Leu Gly Thr Tyr Trp Asp
 110 115 120
 Asn Cys Asn Arg Cys Thr Cys Gln Glu Asn Arg Gln Trp Gln Cys
 125 130 135
 Asp Gln Glu Pro Cys Leu Val Asp Pro Asp Met Ile Lys Ala Ile
 140 145 150

Asn Gln Gly Asn Tyr Gly Trp Gln Ala Gly Asn His Ser Ala Phe
 155 160 165
 Trp Gly Met Thr Leu Asp Glu Gly Ile Arg Tyr Arg Leu Gly Thr
 170 175 180
 Ile Arg Pro Ser Ser Ser Val Met Asn Met His Glu Ile Tyr Thr
 185 190 195
 Val Leu Asn Pro Gly Glu Val Leu Pro Thr Ala Phe Glu Ala Ser
 200 205 210
 Glu Lys Trp Pro Asn Leu Ile His Glu Pro Leu Asp Gln Gly Asn
 215 220 225
 Cys Ala Gly Ser Trp Ala Phe Ser Thr Ala Ala Val Ala Ser Asp
 230 235 240
 Arg Val Ser Ile His Ser Leu Gly His Met Thr Pro Val Leu Ser
 245 250 255
 Pro Gln Asn Leu Leu Ser Cys Asp Thr His Gln Gln Gln Gly Cys
 260 265 270
 Arg Gly Gly Arg Leu Asp Gly Ala Trp Trp Phe Leu Arg Arg Arg
 275 280 285
 Gly Val Val Ser Asp His Cys Tyr Pro Phe Ser Gly Arg Glu Arg
 290 295 300
 Asp Glu Ala Gly Pro Ala Pro Pro Cys Met Met His Ser Arg Ala
 305 310 315
 Met Gly Arg Gly Lys Arg Gln Ala Thr Ala His Cys Pro Asn Ser
 320 325 330
 Tyr Val Asn Asn Asn Asp Ile Tyr Gln Val Thr Pro Val Tyr Arg
 335 340 345
 Leu Gly Ser Asn Asp Lys Glu Ile Met Lys Glu Leu Met Glu Asn
 350 355 360
 Gly Pro Val Gln Ala Leu Met Glu Val His Glu Asp Phe Phe Leu
 365 370 375
 Tyr Lys Gly Gly Ile Tyr Ser His Thr Pro Val Ser Leu Gly Arg
 380 385 390
 Pro Glu Arg Tyr Arg Arg His Gly Thr His Ser Val Lys Ile Thr
 395 400 405
 Gly Trp Gly Glu Glu Thr Leu Pro Asp Gly Arg Thr Leu Lys Tyr
 410 415 420
 Trp Thr Ala Ala Asn Ser Trp Gly Pro Ala Trp Gly Glu Arg Gly
 425 430 435
 His Phe Arg Ile Val Arg Gly Val Asn Glu Cys Asp Ile Glu Ser
 440 445 450
 Phe Val Leu Gly Val Trp Gly Arg Val Gly Met Glu Asp Met Gly
 455 460 465
 His His

<210> 4
 <211> 187
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2769104CD1

<400> 4
 Met Pro Gly Pro Arg Val Trp Gly Lys Tyr Leu Trp Arg Ser Pro
 1 5 10 15
 His Ser Lys Gly Cys Pro Gly Ala Met Trp Trp Leu Leu Leu Trp
 20 25 30
 Gly Val Leu Gln Ala Cys Pro Thr Arg Gly Ser Val Leu Leu Ala
 35 40 45
 Gln Glu Leu Pro Gln Gln Leu Thr Ser Pro Gly Tyr Pro Glu Pro
 50 55 60
 Tyr Gly Lys Gly Gln Glu Ser Ser Thr Asp Ile Lys Ala Pro Glu
 65 70 75
 Gly Phe Ala Val Arg Leu Val Phe Gln Asp Phe Asp Leu Glu Pro
 80 85 90

Ser Gln Asp Cys Ala Gly Asp Ser Val Thr Ile Ser Phe Val Gly
 95 100 105
 Ser Asp Pro Ser Gln Phe Cys Gly Gln Gly Ser Pro Leu Gly
 110 115 120
 Arg Pro Pro Gly Gln Arg Glu Phe Val Ser Ser Gly Arg Ser Leu
 125 130 135
 Arg Leu Thr Phe Arg Thr Gln Pro Ser Ser Glu Asn Lys Thr Ala
 140 145 150
 His Leu His Lys Gly Phe Leu Ala Leu Tyr Gln Thr Val Gly Glu
 155 160 165
 Cys Pro Ser Trp Gly Cys Arg Glu Gly Ala Ser Val Pro Ser His
 170 175 180
 Asp Pro Gly Ile Phe Lys Pro
 185

<210> 5
 <211> 289
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4802789CD1

<400> 5
 Met Arg Val Lys Asp Pro Thr Lys Ala Leu Pro Glu Lys Ala Lys
 1 5 10 15
 Arg Ser Lys Arg Pro Thr Val Pro His Asp Glu Asp Ser Ser Asp
 20 25 30
 Asp Ile Ala Val Gly Leu Thr Cys Gln His Val Ser His Ala Ile
 35 40 45
 Ser Val Asn His Val Lys Arg Ala Ile Ala Glu Asn Leu Trp Ser
 50 55 60
 Val Cys Ser Glu Cys Leu Lys Glu Arg Arg Phe Tyr Asp Gly Gln
 65 70 75
 Leu Val Leu Thr Ser Asp Ile Trp Leu Cys Leu Lys Cys Gly Phe
 80 85 90
 Gln Gly Cys Gly Lys Asn Ser Glu Ser Gln His Ser Leu Lys His
 95 100 105
 Phe Lys Ser Ser Arg Thr Glu Pro His Cys Ile Ile Ile Asn Leu
 110 115 120
 Ser Thr Trp Ile Ile Trp Cys Tyr Glu Cys Asp Glu Lys Leu Ser
 125 130 135
 Thr His Cys Asn Lys Lys Val Leu Ala Gln Ile Val Asp Phe Leu
 140 145 150
 Gln Lys His Ala Ser Lys Thr Gln Thr Ser Ala Phe Ser Arg Ile
 155 160 165
 Met Lys Leu Cys Glu Glu Lys Cys Glu Thr Asp Glu Ile Gln Lys
 170 175 180
 Gly Gly Lys Cys Arg Asn Leu Ser Val Arg Gly Ile Thr Asn Leu
 185 190 195
 Gly Asn Thr Cys Phe Phe Asn Ala Val Met Gln Asn Leu Ala Gln
 200 205 210
 Thr Tyr Thr Leu Thr Asp Leu Met Asn Glu Ile Lys Glu Ser Ser
 215 220 225
 Thr Lys Leu Lys Ile Phe Pro Ser Ser Asp Ser Gln Leu Asp Pro
 230 235 240
 Leu Val Val Glu Leu Ser Arg Pro Gly Pro Leu Thr Ser Ala Leu
 245 250 255
 Phe Leu Phe Leu His Ser Met Lys Glu Thr Glu Lys Gly Pro Leu
 260 265 270
 Ser Pro Lys Val Leu Phe Asn Gln Leu Cys Gln Lys Trp Val His
 275 280 285
 Leu His Leu Ile

<210> 6

<211> 960
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 60116897CD1

<400> 6
Met Phe His Ser Ser Ala Met Val Asn Ser His Arg Lys Pro Met
1 5 10 15
Phe Asn Ile His Arg Gly Phe Tyr Cys Leu Thr Ala Ile Leu Pro
20 25 30
Gln Ile Cys Ile Cys Ser Gln Phe Ser Val Pro Ser Ser Tyr His
35 40 45
Phe Thr Glu Asp Pro Gly Ala Phe Pro Val Ala Thr Asn Gly Glu
50 55 60
Arg Phe Pro Trp Gln Glu Leu Arg Leu Pro Ser Val Val Ile Pro
65 70 75
Leu His Tyr Asp Leu Phe Val His Pro Asn Leu Thr Ser Leu Asp
80 85 90
Phe Val Ala Ser Glu Lys Ile Glu Val Leu Val Ser Asn Ala Thr
95 100 105
Gln Phe Ile Ile Leu His Ser Lys Asp Leu Glu Ile Thr Asn Ala
110 115 120
Thr Leu Gln Ser Glu Glu Asp Ser Arg Tyr Met Lys Pro Gly Lys
125 130 135
Glu Leu Lys Val Leu Ser Tyr Pro Ala His Glu Gln Ile Ala Leu
140 145 150
Leu Val Pro Glu Lys Leu Thr Pro His Leu Lys Tyr Tyr Val Ala
155 160 165
Met Asp Phe Gln Ala Lys Leu Gly Asp Gly Phe Glu Gly Phe Tyr
170 175 180
Lys Ser Thr Tyr Arg Thr Leu Gly Gly Glu Thr Arg Ile Leu Ala
185 190 195
Val Thr Asp Phe Glu Pro Thr Gln Ala Arg Met Ala Phe Pro Cys
200 205 210
Phe Asp Glu Pro Leu Phe Lys Ala Asn Phe Ser Ile Lys Ile Arg
215 220 225
Arg Glu Ser Arg His Ile Ala Leu Ser Asn Met Pro Lys Val Lys
230 235 240
Thr Ile Glu Leu Glu Gly Gly Leu Leu Glu Asp His Phe Glu Thr
245 250 255
Thr Val Lys Met Ser Thr Tyr Leu Val Ala Tyr Ile Val Cys Asp
260 265 270
Phe His Ser Leu Ser Gly Phe Thr Ser Ser Gly Val Lys Val Ser
275 280 285
Ile Tyr Ala Ser Pro Asp Lys Arg Asn Gln Thr His Tyr Ala Leu
290 295 300
Gln Ala Ser Leu Lys Leu Leu Asp Phe Tyr Glu Lys Tyr Phe Asp
305 310 315
Ile Tyr Tyr Pro Leu Ser Lys Leu Asp Leu Ile Ala Ile Pro Asp
320 325 330
Phe Ala Pro Gly Ala Met Glu Asn Trp Gly Leu Ile Thr Tyr Arg
335 340 345
Glu Thr Ser Leu Leu Phe Asp Pro Lys Thr Ser Ser Ala Ser Asp
350 355 360
Lys Leu Trp Val Thr Arg Val Ile Ala His Glu Leu Ala His Gln
365 370 375
Trp Phe Gly Asn Leu Val Thr Met Glu Trp Trp Asn Asp Ile Trp
380 385 390
Leu Lys Glu Gly Phe Ala Lys Tyr Met Glu Leu Ile Ala Val Asn
395 400 405
Ala Thr Tyr Pro Glu Leu Gln Phe Asp Asp Tyr Phe Leu Asn Val
410 415 420
Cys Phe Glu Val Ile Thr Lys Asp Ser Leu Asn Ser Ser Arg Pro
425 430 435

Ile Ser Lys Pro Ala Glu Thr Pro Thr Gln Ile Gln Glu Met Phe
 440 445 450
 Asp Glu Val Ser Tyr Asn Lys Gly Ala Cys Ile Leu Asn Met Leu
 455 460 465
 Lys Asp Phe Leu Gly Glu Glu Lys Phe Gln Lys Gly Ile Ile Gln
 470 475 480
 Tyr Leu Lys Lys Phe Ser Tyr Arg Asn Ala Lys Asn Asp Asp Leu
 485 490 495
 Trp Ser Ser Leu Ser Asn Ser Cys Leu Glu Ser Asp Phe Thr Ser
 500 505 510
 Gly Gly Val Cys His Ser Asp Pro Lys Met Thr Ser Asn Met Leu
 515 520 525
 Ala Phe Leu Gly Glu Asn Ala Glu Val Lys Glu Met Met Thr Thr
 530 535 540
 Trp Thr Leu Gln Lys Gly Ile Pro Leu Leu Val Val Lys Gln Asp
 545 550 555
 Gly Cys Ser Leu Arg Leu Gln Gln Glu Arg Phe Leu Gln Gly Val
 560 565 570
 Phe Gln Glu Asp Pro Glu Trp Arg Ala Leu Gln Glu Arg Tyr Leu
 575 580 585
 Trp His Ile Pro Leu Thr Tyr Ser Thr Ser Ser Asn Val Ile
 590 595 600
 His Arg His Ile Leu Lys Ser Lys Thr Asp Thr Leu Asp Leu Pro
 605 610 615
 Glu Lys Thr Ser Trp Val Lys Phe Asn Val Asp Ser Asn Gly Tyr
 620 625 630
 Tyr Ile Val His Tyr Glu Gly His Gly Trp Asp Gln Leu Ile Thr
 635 640 645
 Gln Leu Asn Gln Asn His Thr Leu Leu Arg Pro Lys Asp Arg Val
 650 655 660
 Gly Leu Ile His Asp Val Phe Gln Leu Val Gly Ala Gly Arg Leu
 665 670 675
 Thr Leu Asp Lys Ala Leu Asp Met Thr Tyr Tyr Leu Gln His Glu
 680 685 690
 Thr Ser Ser Pro Ala Leu Leu Glu Gly Leu Ser Tyr Leu Glu Ser
 695 700 705
 Phe Tyr His Met Met Asp Arg Arg Asn Ile Ser Asp Ile Ser Glu
 710 715 720
 Asn Leu Lys Arg Tyr Leu Leu Gln Tyr Phe Lys Pro Val Ile Asp
 725 730 735
 Arg Gln Ser Trp Ser Asp Lys Gly Ser Val Trp Asp Arg Met Leu
 740 745 750
 Arg Ser Ala Leu Leu Lys Leu Ala Cys Asp Leu Asn His Ala Pro
 755 760 765
 Cys Ile Gln Lys Ala Ala Glu Leu Phe Ser Gln Trp Met Glu Ser
 770 775 780
 Ser Gly Lys Leu Asn Ile Pro Thr Asp Val Leu Lys Ile Val Tyr
 785 790 795
 Ser Val Gly Ala Gln Thr Thr Ala Gly Trp Asn Tyr Leu Leu Glu
 800 805 810
 Gln Tyr Glu Leu Ser Met Ser Ser Ala Glu Gln Asn Lys Ile Leu
 815 820 825
 Tyr Ala Leu Ser Thr Ser Lys His Gln Glu Lys Leu Leu Lys Leu
 830 835 840
 Ile Glu Leu Gly Met Glu Gly Lys Val Ile Lys Thr Gln Asn Leu
 845 850 855
 Ala Ala Leu Leu His Ala Ile Ala Arg Arg Pro Lys Gly Gln Gln
 860 865 870
 Leu Ala Trp Asp Phe Val Arg Glu Asn Trp Thr His Leu Leu Lys
 875 880 885
 Lys Phe Asp Leu Gly Ser Tyr Asp Ile Arg Met Ile Ile Ser Gly
 890 895 900
 Thr Thr Ala His Phe Ser Ser Lys Asp Lys Leu Gln Glu Val Lys
 905 910 915
 Leu Phe Phe Glu Ser Leu Glu Ala Gln Gly Ser His Leu Asp Ile
 920 925 930
 Phe Gln Thr Val Leu Glu Thr Ile Thr Lys Asn Ile Lys Trp Leu

935	940	945
Glu Lys Asn Leu Pro Thr Leu Arg Thr Trp	Leu Met Val Asn Thr	
950	955	960

<210> 7
 <211> 525
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1866356CD1

<400> 7
 Met Ala Val Pro Gly Glu Ala Glu Glu Ala Thr Val Tyr Leu
 1 5 10 15
 Val Val Ser Gly Ile Pro Ser Val Leu Arg Ser Ala His Leu Arg
 20 25 30
 Ser Tyr Phe Ser Gln Phe Arg Glu Glu Arg Gly Gly Gly Phe Leu
 35 40 45
 Cys Phe His Tyr Arg His Arg Pro Glu Arg Ala Pro Pro Gln Ala
 50 55 60
 Ala Pro Asn Ser Ala Leu Ile Pro Thr Asp Pro Ala Ala Glu Gly
 65 70 75
 Gln Leu Leu Ser Gln Thr Ser Ala Thr Asp Val Arg Pro Leu Ser
 80 85 90
 Thr Arg Asp Ser Thr Pro Ile Gln Thr Arg Thr Cys Cys Cys Val
 95 100 105
 Ile Ser Val Arg Gly Leu Ala Gln Ala Gln Arg Leu Ile Arg Met
 110 115 120
 Tyr Ser Gly Arg Arg Trp Leu Asp Ser His Gly Thr Trp Leu Pro
 125 130 135
 Gly Arg Cys Leu Ile Arg Arg Leu Arg Leu Pro Thr Glu Ala Ser
 140 145 150
 Gly Leu Gly Ser Phe Pro Phe Lys Thr Arg Lys Glu Leu Gln Ser
 155 160 165
 Trp Lys Ala Glu Asn Glu Ala Phe Thr Leu Ala Asp Leu Lys Gln
 170 175 180
 Leu Pro Glu Leu Asn Pro Pro Val Leu Met Pro Arg Gly Asn Val
 185 190 195
 Gly Thr Pro Leu Arg Val Phe Leu Glu Leu Ile Arg Ala Cys Arg
 200 205 210
 Leu Pro Pro Arg Ile Ile Thr Gln Leu Gln Leu Gln Phe Pro Lys
 215 220 225
 Thr Gly Ser Ser Arg Arg Tyr Gly Asn Val Pro Phe Glu Tyr Glu
 230 235 240
 Asp Ser Glu Thr Val Glu Gln Glu Glu Leu Val Tyr Thr Ala Glu
 245 250 255
 Gly Glu Glu Ile Pro Gln Gly Thr Tyr Leu Ala Asp Ile Pro Ala
 260 265 270
 Ser Pro Cys Gly Glu Pro Glu Glu Glu Val Gly Lys Glu Glu Glu
 275 280 285
 Glu Glu Ser His Ser Asp Glu Asp Asp Asp Arg Gly Glu Glu Trp
 290 295 300
 Glu Arg His Glu Ala Leu His Glu Asp Val Thr Gly Gln Glu Arg
 305 310 315
 Thr Thr Glu Gln Leu Phe Glu Glu Glu Ile Glu Leu Lys Trp Glu
 320 325 330
 Lys Gly Gly Ser Gly Leu Val Phe Tyr Thr Asp Ala Gln Phe Trp
 335 340 345
 Gln Glu Glu Glu Gly Asp Phe Asp Glu Gln Thr Ala Asp Asp Trp
 350 355 360
 Asp Val Asp Met Ser Val Tyr Tyr Asp Arg Asp Gly Gly Asp Lys
 365 370 375
 Asp Ala Arg Asp Ser Val Gln Met Arg Leu Glu Gln Arg Leu Arg
 380 385 390

Asp	Gly	Gln	Glu	Asp	Gly	Ser	Val	Ile	Glu	Arg	Gln	Val	Gly	Thr
				395					400					405
Phe	Glu	Arg	His	Thr	Lys	Gly	Ile	Gly	Arg	Lys	Val	Met	Glu	Arg
				410					415					420
Gln	Gly	Trp	Ala	Glu	Gly	Gln	Gly	Leu	Gly	Cys	Arg	Cys	Ser	Gly
				425					430					435
Val	Pro	Glu	Ala	Leu	Asp	Ser	Asp	Gly	Gln	His	Pro	Arg	Cys	Lys
				440					445					450
Arg	Gly	Leu	Gly	Tyr	His	Gly	Glu	Lys	Leu	Gln	Pro	Phe	Gly	Gln
				455					460					465
Leu	Lys	Arg	Pro	Arg	Arg	Asn	Gly	Leu	Gly	Leu	Ile	Ser	Thr	Ile
				470					475					480
Tyr	Asp	Glu	Pro	Leu	Pro	Gln	Asp	Gln	Thr	Glu	Ser	Leu	Leu	Arg
				485					490					495
Arg	Gln	Pro	Pro	Thr	Ser	Met	Lys	Phe	Arg	Thr	Asp	Met	Ala	Phe
				500					505					510
Val	Arg	Gly	Ser	Ser	Cys	Ala	Ser	Asp	Ser	Pro	Ser	Leu	Pro	Asp
				515					520					525

<210> 8
<211> 795
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1872095CD1

<400> 8															
Met	Ile	Thr	Val	Leu	Ile	Arg	Ser	Leu	Thr	Thr	Asp	Pro	Asn	Val	
1				5					10					15	
Lys	Asp	Ala	Ser	Met	Thr	Gln	Ala	Leu	Cys	Arg	Met	Ile	Asp	Trp	
				20					25					30	
Leu	Ser	Trp	Pro	Leu	Ala	Gln	His	Val	Asp	Thr	Trp	Val	Ile	Ala	
				35					40					45	
Leu	Leu	Lys	Gly	Leu	Ala	Ala	Val	Gln	Lys	Phe	Thr	Ile	Leu	Ile	
				50					55					60	
Asp	Val	Thr	Leu	Leu	Lys	Ile	Glu	Leu	Val	Phe	Asn	Arg	Leu	Trp	
				65					70					75	
Phe	Pro	Leu	Val	Arg	Pro	Gly	Ala	Leu	Ala	Val	Leu	Ser	His	Met	
				80					85					90	
Leu	Leu	Ser	Phe	Gln	His	Ser	Pro	Glu	Ala	Phe	His	Leu	Ile	Val	
				95					100					105	
Pro	His	Val	Val	Asn	Leu	Val	His	Ser	Phe	Lys	Asn	Asp	Gly	Leu	
				110					115					120	
Pro	Ser	Ser	Thr	Ala	Phe	Leu	Val	Gln	Leu	Thr	Glu	Leu	Ile	His	
				125					130					135	
Cys	Met	Met	Tyr	His	Tyr	Ser	Gly	Phe	Pro	Asp	Leu	Tyr	Glu	Pro	
				140					145					150	
Ile	Leu	Glu	Ala	Ile	Lys	Asp	Phe	Pro	Lys	Pro	Ser	Glu	Glu	Lys	
				155					160					165	
Ile	Lys	Leu	Ile	Leu	Asn	Gln	Ser	Ala	Trp	Thr	Ser	Gln	Ser	Asn	
				170					175					180	
Ser	Leu	Ala	Ser	Cys	Leu	Ser	Arg	Leu	Ser	Gly	Lys	Ser	Glu	Thr	
				185					190					195	
Gly	Lys	Thr	Gly	Leu	Ile	Asn	Leu	Gly	Asn	Thr	Cys	Tyr	Met	Asn	
				200					205					210	
Ser	Val	Ile	Gln	Ala	Leu	Phe	Met	Ala	Thr	Asp	Phe	Arg	Arg	Gln	
				215					220					225	
Val	Leu	Ser	Leu	Asn	Leu	Asn	Gly	Cys	Asn	Ser	Leu	Met	Lys	Lys	
				230					235					240	
Leu	Gln	His	Leu	Phe	Ala	Phe	Leu	Ala	His	Thr	Gln	Arg	Glu	Ala	
				245					250					255	
Tyr	Ala	Pro	Arg	Ile	Phe	Phe	Glu	Ala	Ser	Arg	Pro	Pro	Trp	Phe	
				260					265					270	
Thr	Pro	Arg	Ser	Gln	Gln	Asp	Cys	Ser	Glu	Tyr	Leu	Arg	Phe	Leu	

275	280	285
Leu Asp Arg Leu His Glu Glu Glu Lys Ile Leu Lys Val Gln Ala		
290	295	300
Ser His Lys Pro Ser Glu Ile Leu Glu Cys Ser Glu Thr Ser Leu		
305	310	315
Gln Glu Val Ala Ser Lys Ala Ala Val Leu Thr Glu Thr Pro Arg		
320	325	330
Thr Ser Asp Gly Glu Lys Thr Leu Ile Glu Lys Met Phe Gly Gly		
335	340	345
Lys Leu Arg Thr His Ile Arg Cys Leu Asn Cys Arg Ser Thr Ser		
350	355	360
Gln Lys Val Glu Ala Phe Thr Asp Leu Ser Leu Ala Phe Cys Pro		
365	370	375
Ser Ser Ser Leu Glu Asn Met Ser Val Gln Asp Pro Ala Ser Ser		
380	385	390
Pro Ser Ile Gln Asp Gly Gly Leu Met Gln Ala Ser Val Pro Gly		
395	400	405
Pro Ser Glu Glu Pro Val Val Tyr Asn Pro Thr Thr Ala Ala Phe		
410	415	420
Ile Cys Asp Ser Leu Val Asn Glu Lys Thr Ile Gly Ser Pro Pro		
425	430	435
Asn Glu Phe Tyr Cys Ser Glu Asn Thr Ser Val Pro Asn Glu Ser		
440	445	450
Asn Lys Ile Leu Val Asn Lys Asp Val Pro Gln Lys Pro Gly Gly		
455	460	465
Glu Thr Thr Pro Ser Val Thr Asp Leu Leu Asn Tyr Phe Leu Ala		
470	475	480
Pro Glu Ile Leu Thr Gly Asp Asn Gln Tyr Tyr Cys Glu Asn Cys		
485	490	495
Ala Ser Leu Gln Asn Ala Glu Lys Thr Met Gln Ile Thr Glu Glu		
500	505	510
Pro Glu Tyr Leu Ile Leu Thr Leu Leu Arg Phe Ser Tyr Asp Gln		
515	520	525
Lys Tyr His Val Arg Arg Lys Ile Leu Asp Asn Val Ser Leu Pro		
530	535	540
Leu Val Leu Glu Leu Pro Val Lys Arg Ile Thr Ser Phe Ser Ser		
545	550	555
Leu Ser Glu Ser Trp Ser Val Asp Val Asp Phe Thr Asp Leu Ser		
560	565	570
Glu Asn Leu Ala Lys Lys Leu Lys Pro Ser Gly Thr Asp Glu Ala		
575	580	585
Ser Cys Thr Lys Leu Val Pro Tyr Leu Leu Ser Ser Val Val Val		
590	595	600
His Ser Gly Ile Ser Ser Glu Ser Gly His Tyr Tyr Ser Tyr Ala		
605	610	615
Arg Asn Ile Thr Ser Thr Asp Ser Ser Tyr Gln Met Tyr His Gln		
620	625	630
Ser Glu Ala Leu Ala Leu Ala Ser Ser Gln Ser His Leu Leu Gly		
635	640	645
Arg Asp Ser Pro Ser Ala Val Phe Glu Gln Asp Leu Glu Asn Lys		
650	655	660
Glu Met Ser Lys Glu Trp Phe Leu Phe Asn Asp Ser Arg Val Thr		
665	670	675
Phe Thr Ser Phe Gln Ser Val Gln Lys Ile Thr Ser Arg Phe Pro		
680	685	690
Lys Asp Thr Ala Tyr Val Leu Leu Tyr Lys Lys Gln His Ser Thr		
695	700	705
Asn Gly Leu Ser Gly Asn Asn Pro Thr Ser Gly Leu Trp Ile Asn		
710	715	720
Gly Asp Pro Pro Leu Gln Lys Glu Leu Met Asp Ala Ile Thr Lys		
725	730	735
Asp Asn Lys Leu Tyr Leu Gln Glu Gln Glu Leu Asn Ala Arg Ala		
740	745	750
Arg Ala Leu Gln Ala Ala Ser Ala Ser Cys Ser Phe Arg Pro Asn		
755	760	765
Gly Phe Asp Asp Asn Asp Pro Pro Gly Ser Cys Gly Pro Thr Gly		
770	775	780

Gly Gly Gly Gly Gly Phe Asn Thr Val Gly Arg Leu Val Phe
 785 790 795

<210> 9
 <211> 919
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2278688CD1

<400> 9
 Met Trp Leu Ala Ala Ala Ala Pro Ser Leu Ala Arg Arg Leu Leu
 1 5 10 15
 Phe Leu Gly Pro Pro Pro Pro Pro Leu Leu Leu Val Phe Ser
 20 25 30
 Arg Ser Ser Arg Arg Arg Leu His Ser Leu Gly Leu Ala Ala Met
 35 40 45
 Pro Glu Lys Arg Pro Phe Glu Arg Leu Pro Ala Asp Val Ser Pro
 50 55 60
 Ile Asn Cys Ser Leu Cys Leu Lys Pro Asp Leu Leu Asp Phe Thr
 65 70 75
 Phe Glu Gly Lys Leu Glu Ala Ala Gln Val Arg Gln Ala Thr
 80 85 90
 Asn Gln Ile Val Met Asn Cys Ala Asp Ile Asp Ile Ile Thr Ala
 95 100 105
 Ser Tyr Ala Pro Glu Gly Asp Glu Glu Ile His Ala Thr Gly Phe
 110 115 120
 Asn Tyr Gln Asn Glu Asp Glu Lys Val Thr Leu Ser Phe Pro Ser
 125 130 135
 Thr Leu Gln Thr Gly Thr Gly Thr Leu Lys Ile Asp Phe Val Gly
 140 145 150
 Glu Leu Asn Asp Lys Met Lys Gly Phe Tyr Arg Ser Lys Tyr Thr
 155 160 165
 Thr Pro Ser Gly Glu Val Arg Tyr Ala Ala Val Thr Gln Phe Glu
 170 175 180
 Ala Thr Asp Ala Arg Arg Ala Phe Pro Cys Trp Asp Glu Pro Ala
 185 190 195
 Ile Lys Ala Thr Phe Asp Ile Ser Leu Val Val Pro Lys Asp Arg
 200 205 210
 Val Ala Leu Ser Asn Met Asn Val Ile Asp Arg Lys Pro Tyr Pro
 215 220 225
 Asp Asp Glu Asn Leu Val Glu Val Lys Phe Ala Arg Thr Pro Val
 230 235 240
 Met Ser Thr Tyr Leu Val Ala Phe Val Val Gly Glu Tyr Asp Phe
 245 250 255
 Val Glu Thr Arg Ser Lys Asp Gly Val Cys Val Arg Val Tyr Thr
 260 265 270
 Pro Val Gly Lys Ala Glu Gln Gly Lys Phe Ala Leu Glu Val Ala
 275 280 285
 Ala Lys Thr Leu Pro Phe Tyr Lys Asp Tyr Phe Asn Val Pro Tyr
 290 295 300
 Pro Leu Pro Lys Ile Asp Leu Ile Ala Ile Ala Asp Phe Ala Ala
 305 310 315
 Gly Ala Met Glu Asn Trp Gly Leu Val Thr Tyr Arg Glu Thr Ala
 320 325 330
 Leu Leu Ile Asp Pro Lys Asn Ser Cys Ser Ser Ser Arg Gln Trp
 335 340 345
 Val Ala Leu Val Val Gly His Glu Leu Ala His Gln Trp Phe Gly
 350 355 360
 Asn Leu Val Thr Met Glu Trp Trp Thr His Leu Trp Leu Asn Glu
 365 370 375
 Gly Phe Ala Ser Trp Ile Glu Tyr Leu Cys Val Asp His Cys Phe
 380 385 390
 Pro Glu Tyr Asp Ile Trp Thr Gln Phe Val Ser Ala Asp Tyr Thr

395	400	405
Arg Ala Gln Glu	Leu Asp Ala Leu Asp	Asn Ser His Pro Ile
410	415	420
Val Ser Val Gly	His Pro Ser Glu Val	Asp Glu Ile Phe Asp
425	430	435
Ile Ser Tyr Ser	Lys Gly Ala Ser Val	Ile Arg Met Leu His
440	445	450
Tyr Ile Gly Asp	Lys Asp Phe Lys	Gly Met Asn Met Tyr
455	460	465
Thr Lys Phe Gln	Gln Lys Asn Ala Ala	Thr Glu Asp Leu Trp
470	475	480
Ser Leu Glu Asn	Ala Ser Gly Lys Pro	Ile Ala Ala Val Met
485	490	495
Thr Trp Thr Lys	Gln Met Gly Phe Pro	Leu Ile Tyr Val Glu
500	505	510
Glu Gln Val Glu	Asp Asp Arg Leu Leu	Arg Leu Ser Gln Lys
515	520	525
Phe Cys Ala Gly	Gly Ser Tyr Val Gly	Glu Asp Cys Pro Gln
530	535	540
Met Val Pro Ile	Thr Ile Ser Thr Ser	Glu Asp Pro Asn Gln
545	550	555
Lys Leu Lys Ile	Leu Met Asp Lys Pro	Glu Met Asn Val Val
560	565	570
Lys Asn Val Lys	Pro Asp Gln Trp Val	Lys Leu Asn Leu Gly
575	580	585
Val Gly Phe Tyr	Arg Thr Gln Tyr Ser	Ser Ala Met Leu Glu
590	595	600
Leu Leu Pro Gly	Ile Arg Asp Leu Ser	Leu Pro Pro Val Asp
605	610	615
Leu Gly Leu Gln	Asn Asp Leu Phe Ser	Leu Ala Arg Ala Gly
620	625	630
Ile Ser Thr Val	Glu Val Leu Lys Val	Met Glu Ala Phe Val
635	640	645
Glu Pro Asn Tyr	Thr Val Trp Ser Asp	Leu Ser Cys Asn Leu
650	655	660
Ile Leu Ser Thr	Leu Leu Ser His Thr	Asp Phe Tyr Glu Glu
665	670	675
Gln Glu Phe Val	Lys Asp Val Phe Ser	Pro Ile Gly Glu Arg
680	685	690
Gly Trp Asp Pro	Lys Pro Gly Glu Gly	His Leu Asp Ala Leu
695	700	705
Arg Gly Leu Val	Leu Gly Lys Leu Gly	Lys Ala Gly His Lys
710	715	720
Thr Leu Glu Glu	Ala Arg Arg Arg Phe	Lys Asp His Val Glu
725	730	735
Lys Gln Ile Leu	Ser Ala Asp Leu Arg	Ser Pro Val Tyr Leu
740	745	750
Val Leu Lys His	Gly Asp Gly Thr Thr	Leu Asp Ile Met Leu
755	760	765
Leu His Lys Gln	Ala Asp Met Gln Glu	Glu Lys Asn Arg Ile
770	775	780
Arg Val Leu Gly	Ala Thr Leu Leu Pro	Asp Leu Ile Gln Lys
785	790	795
Leu Thr Phe Ala	Leu Ser Glu Glu Val	Arg Pro Gln Asp Thr
800	805	810
Ser Val Ile Gly	Gly Val Ala Gly Gly	Ser Lys His Gly Arg
815	820	825
Ala Ala Trp Lys	Phe Ile Lys Asp Asn	Trp Glu Glu Leu Tyr
830	835	840
Arg Tyr Gln Gly	Gly Phe Leu Ile Ser	Arg Leu Ile Lys Leu
845	850	855
Val Glu Gly Phe	Ala Val Asp Lys Met	Ala Gly Glu Val Lys
860	865	870
Phe Phe Glu Ser	His Pro Ala Pro Ser	Ala Glu Arg Thr Ile
875	880	885
Gln Cys Cys Glu	Asn Ile Leu Leu Asn	Ala Ala Trp Leu Lys
890	895	900

Asp Ala Glu Ser Ile His Gln Tyr Leu Leu Gln Arg Lys Ala Ser
 905 910 915
 Pro Pro Thr Val

<210> 10
 <211> 209
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4043361CD1

<400> 10
 Met Glu Gln Pro Arg Lys Ala Val Val Val Thr Gly Phe Gly Pro
 1 5 10 15
 Phe Gly Glu His Thr Val Asn Ala Ser Trp Ile Ala Val Gln Glu
 20 25 30
 Leu Glu Lys Leu Gly Leu Gly Asp Ser Val Asp Leu His Val Tyr
 35 40 45
 Glu Ile Pro Val Glu Tyr Gln Thr Val Gln Arg Leu Ile Pro Ala
 50 55 60
 Leu Trp Glu Lys His Ser Pro Gln Leu Val Val His Val Gly Val
 65 70 75
 Ser Gly Met Ala Thr Thr Val Thr Leu Glu Lys Cys Gly His Asn
 80 85 90
 Lys Gly Tyr Lys Gly Leu Asp Asn Cys Arg Phe Cys Pro Gly Ser
 95 100 105
 Gln Cys Cys Val Glu Asp Gly Pro Glu Ser Ile Asp Ser Ile Ile
 110 115 120
 Asp Met Asp Ala Val Cys Lys Arg Val Thr Thr Leu Gly Leu Asp
 125 130 135
 Val Ser Val Thr Ile Ser Gln Asp Ala Gly Arg Lys Lys Pro Phe
 140 145 150
 Pro Ala Lys Gly Asp Cys Val Phe Cys Arg Arg Arg Arg Ala Arg
 155 160 165
 Ser Leu Gln Ala Gln Cys Gly Phe Ser Leu Thr Pro Ala Leu Glu
 170 175 180
 Leu Leu Pro Val Pro Phe Leu Lys Leu Leu Cys Pro Gly Pro Pro
 185 190 195
 Arg Arg Arg Arg Ile Cys Arg Ile Leu Pro Gly Ala Gly Leu
 200 205

<210> 11
 <211> 77
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3937958CD1

<400> 11
 Met Gly Lys Glu Lys Ala Leu Ser Leu Gln Met Met Lys Tyr Trp
 1 5 10 15
 Ala Asn Phe Ala Arg Thr Gly Asn Pro Asn Asp Gly Asn Leu Pro
 20 25 30
 Cys Trp Pro Arg Tyr Asn Lys Asp Glu Lys Tyr Leu Gln Leu Asp
 35 40 45
 Phe Thr Thr Arg Val Gly Met Lys Leu Lys Glu Lys Lys Met Ala
 50 55 60
 Phe Trp Met Ser Leu Tyr Gln Ser Gln Arg Pro Glu Lys Gln Arg
 65 70 75
 Gln Phe

<210> 12
 <211> 414
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7257324CD1

<400> 12
 Met Asn Pro Thr Leu Gly Leu Ala Ile Phe Leu Ala Val Leu Leu
 1 5 10 15
 Thr Val Lys Gly Leu Leu Lys Pro Ser Phe Ser Pro Arg Asn Tyr
 20 25 30
 Lys Ala Leu Ser Glu Val Gln Gly Trp Lys Gln Arg Met Ala Ala
 35 40 45
 Lys Glu Leu Ala Arg Gln Asn Met Asp Leu Gly Phe Lys Leu Leu
 50 55 60
 Lys Lys Leu Ala Phe Tyr Asn Pro Gly Arg Asn Ile Phe Leu Ser
 65 70 75
 Pro Leu Ser Ile Ser Thr Ala Phe Ser Met Leu Cys Leu Gly Ala
 80 85 90
 Gln Asp Ser Thr Leu Asp Glu Ile Lys Gln Gly Phe Asn Phe Arg
 95 100 105
 Lys Met Pro Glu Lys Asp Leu His Glu Gly Phe His Tyr Ile Ile
 110 115 120
 His Glu Leu Thr Gln Lys Thr Gln Asp Leu Lys Leu Ser Ile Gly
 125 130 135
 Asn Thr Leu Phe Ile Asp Gln Arg Leu Gln Pro Gln Arg Lys Phe
 140 145 150
 Leu Glu Asp Ala Lys Asn Phe Tyr Ser Ala Glu Thr Ile Leu Thr
 155 160 165
 Asn Phe Gln Asn Leu Glu Met Ala Gln Lys Gln Ile Asn Asp Phe
 170 175 180
 Ile Ser Gln Lys Thr His Gly Lys Ile Asn Asn Leu Ile Glu Asn
 185 190 195
 Ile Asp Pro Gly Thr Val Met Leu Leu Ala Asn Tyr Ile Phe Phe
 200 205 210
 Arg Ala Arg Trp Lys His Glu Phe Asp Pro Asn Val Thr Lys Glu
 215 220 225
 Glu Asp Phe Phe Leu Glu Lys Asn Ser Ser Val Lys Val Pro Met
 230 235 240
 Met Phe Arg Ser Gly Ile Tyr Gln Val Gly Tyr Asp Asp Lys Leu
 245 250 255
 Ser Cys Thr Ile Leu Glu Ile Pro Tyr Gln Lys Asn Ile Thr Ala
 260 265 270
 Ile Phe Ile Leu Pro Asp Glu Gly Lys Leu Lys His Leu Glu Lys
 275 280 285
 Gly Leu Gln Val Asp Thr Phe Ser Arg Trp Lys Thr Leu Leu Ser
 290 295 300
 Arg Arg Val Val Asp Val Ser Val Pro Arg Leu His Met Thr Gly
 305 310 315
 Thr Phe Asp Leu Lys Lys Thr Leu Ser Tyr Ile Gly Val Ser Lys
 320 325 330
 Ile Phe Glu Glu His Gly Asp Leu Thr Lys Ile Ala Pro His Arg
 335 340 345
 Ser Leu Lys Val Gly Glu Ala Val His Lys Ala Glu Leu Lys Met
 350 355 360
 Asp Glu Arg Gly Thr Glu Gly Ala Ala Gly Thr Gly Ala Gln Thr
 365 370 375
 Leu Pro Met Glu Thr Pro Leu Val Val Lys Ile Asp Lys Pro Tyr
 380 385 390
 Leu Leu Leu Ile Tyr Ser Glu Lys Ile Pro Ser Val Leu Phe Leu
 395 400 405
 Gly Lys Ile Val Asn Pro Ile Gly Lys
 410

<210> 13
 <211> 397
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472038CD1

<400> 13
 Met Pro Arg Ala Ile Ser Pro Leu Met Arg Phe Gln His Pro Val
 1 5 10 15
 Ser Cys Lys Leu Gln Leu Tyr Arg Val Pro Leu Arg Arg Phe Pro
 20 25 30
 Ser Ala Arg His Arg Phe Glu Lys Leu Gly Ile Arg Met Asp Arg
 35 40 45
 Leu Arg Leu Lys Tyr Ala Glu Glu Val Ser His Phe Arg Gly Glu
 50 55 60
 Trp Asn Ser Ala Val Lys Ser Thr Pro Leu Ser Asn Tyr Leu Asp
 65 70 75
 Ala Gln Tyr Phe Gly Pro Ile Thr Ile Gly Thr Pro Pro Gln Thr
 80 85 90
 Phe Lys Val Ile Phe Asp Thr Gly Ser Ser Asn Leu Trp Val Pro
 95 100 105
 Ser Ala Thr Cys Ala Ser Thr Met Val Ala Cys Arg Val His Asn
 110 115 120
 Arg Tyr Phe Ala Lys Arg Ser Thr Ser His Gln Val Arg Gly Asp
 125 130 135
 His Phe Ala Ile His Tyr Gly Ser Gly Ser Leu Ser Gly Phe Leu
 140 145 150
 Ser Thr Asp Thr Val Arg Val Ala Gly Leu Glu Ile Arg Asp Gln
 155 160 165
 Thr Phe Ala Glu Ala Thr Glu Met Pro Gly Pro Ile Phe Leu Ala
 170 175 180
 Ala Lys Phe Asp Gly Ile Phe Gly Leu Ala Tyr Arg Ser Ile Ser
 185 190 195
 Met Gln Arg Ile Lys Pro Pro Phe Tyr Ala Met Met Glu Gln Gly
 200 205 210
 Leu Leu Thr Lys Pro Ile Phe Ser Val Tyr Leu Ser Arg Asn Gly
 215 220 225
 Glu Lys Asp Gly Gly Ala Ile Phe Phe Gly Gly Ser Asn Pro His
 230 235 240
 Tyr Tyr Thr Gly Asn Phe Thr Tyr Val Gln Val Ser His Arg Ala
 245 250 255
 Tyr Trp Gln Val Lys Met Asp Ser Ala Val Ile Arg Asn Leu Glu
 260 265 270
 Leu Cys Gln Gln Gly Cys Glu Val Ile Ile Asp Thr Gly Thr Ser
 275 280 285
 Phe Leu Ala Leu Pro Tyr Asp Gln Ala Ile Leu Ile Asn Glu Ser
 290 295 300
 Ile Gly Gly Thr Pro Ser Ser Phe Gly Gln Phe Leu Val Pro Cys
 305 310 315
 Asp Ser Val Pro Asp Leu Pro Lys Ile Thr Phe Thr Leu Gly Gly
 320 325 330
 Arg Arg Phe Phe Leu Glu Ser His Glu Tyr Val Phe Arg Asp Ile
 335 340 345
 Tyr Gln Asp Arg Arg Ile Cys Ser Ser Ala Phe Ile Ala Val Asp
 350 355 360
 Leu Pro Ser Pro Ser Gly Pro Leu Trp Ile Leu Gly Asp Val Phe
 365 370 375
 Leu Gly Lys Tyr Tyr Thr Glu Phe Asp Met Glu Arg His Arg Ile
 380 385 390
 Gly Phe Ala Asp Ala Arg Ser
 395

<210> 14
 <211> 145

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472041CD1

<400> 14
 Met Gly Ile Gly Cys Trp Arg Asn Pro Leu Leu Leu Leu Ile Ala
 1 5 10 15
 Leu Val Leu Ser Ala Lys Leu Gly His Phe Gln Arg Trp Glu Gly
 20 25 30
 Phe Gln Gln Lys Leu Met Ser Lys Lys Asn Met Asn Ser Thr Leu
 35 40 45
 Asn Phe Phe Ile Gln Ser Tyr Asn Asn Ala Ser Asn Asp Thr Tyr
 50 55 60
 Leu Tyr Arg Val Gln Arg Leu Ile Arg Ser Gln Met Gln Leu Thr
 65 70 75
 Thr Gly Val Glu Tyr Ile Val Thr Val Lys Ile Gly Trp Thr Lys
 80 85 90
 Cys Lys Arg Asn Asp Thr Ser Asn Ser Ser Cys Pro Leu Gln Ser
 95 100 105
 Lys Lys Leu Arg Lys Ser Leu Ile Cys Glu Ser Leu Ile Tyr Thr
 110 115 120
 Met Pro Trp Ile Asn Tyr Phe Gln Leu Trp Asn Asn Ser Cys Leu
 125 130 135
 Glu Ala Glu His Val Gly Arg Asn Leu Arg
 140 145

<210> 15
 <211> 4028
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1714846CB1

<400> 15
 gccattccgg cgccgcgtc cctccggc cctctctccc ttccccaaag cagccccgg 60
 accggcagca aaggaacgtg cgaacgcgtg acgcgcggc actggctcgc gctctccgt 120
 gccccggcgt cctccgccc ctcatggccc gggccgcccggc ggacgagcgg cgctgaggcg 180
 ggccgcgtgg agacgtgagg cggccgcgtgg cgccttcaca gtcggcggtt cgccgcctgc 240
 ccgcggtgcc cgccgcacggc ggccgcgcgc gccttcgcgc ctggctggcg ggggcgtgt 300
 cctcccaaggc cgtccgcgcgc gtcctctggc gtcggcgccg ggcgcgcgc 360
 ggagggcgcga ggagccgggc gccacccggc cgcgcgcgc gcgcgcgcgc 420
 ccgtggagca gaacgtgctg cagcagagcg cggccgcagaa gcaccaggcag acgttttga 480
 atcaacttaga agaaaaattaaatc gggatataatc acacccagat actacagcaa gccttgaagg 540
 atagaatgg aaacttggaa tttagcagtgg ctttccttac tgcgaagaat gctaagaccc 600
 ctcagcagga ggagacacaact tactacaaa cagcacttcc tggcaatgtat agatacatca 660
 gtgtggaaag ccaagcagat acaaattgtga ttgatctcac tggagatgtat aaagatgtatc 720
 ttccatggagc aattgccttg agtttggccg aatcaaacag ggcattcagg gagactggaa 780
 taactgtga ggaacaaggc attagcagag ttcttgaagc cagcatagca gagaataaag 840
 catgtttgaa gaggacacact acagaatgtt ggaggatttc tcgaaacctt tatgtatggaa 900
 aaagacaggc caaaatccc gttggctaa agaatgttg caataactgtt tggttttagt 960
 ctgttattca gtcatttattt aatcttttg aattttagaaat tagttctt aattacaagc 1020
 ctccatcaaa tgctcaagat ttacccggaa accaaaaaggaa acatcgaaat ttgcctttta 1080
 tgcgtgagct gaggtatcta tttgcacttc ttgttggtac caaaaggaaat tatgttgc 1140
 catcaagagc agttgaaatt cttaggatg ctttcaaatttca aatgactca cagcagcaag 1200
 atgtgagtga gtttacacac aaattttagat attgggttaga agatgccttc caaatgaaag 1260
 ctgaagagga gacggatgaa gagaagccaa agaaccat ggttagatgtt ttctatggca 1320
 gatccctggc tttggggatc cttgaaggta aaaaattttttaa aacactgaa atgtttggtc 1380
 agtacccact tcaggtaat gggttcaaaat gttctgcata gttcgtatggaa gtcgtatggaa 1440
 ttgaaggaga aattttagatc ttacatttcg agaattcagg aaaatcaggc caagagcatt 1500
 ggttactga attaccacat gtgttaacat ttgaattgtc aagatttggaa ttatattcagg 1560
 cattggaaag accagaaaaaa attcacaaca aattttagatttcccaagtt ttatatttgg 1620
 acagatacat gcacagaaac agagaaataa caagaattaa gagggaaagag atcaagagac 1680

taaaagatta cctcacggtt ttacaacaaa ggcttagaaag atatthaagc tatggttccg 1740
 gtcccaaacg atccccctt gtagatgtt ctcgtatgc attggattt gcctcaagta 1800
 aacctgtttg cacttctct gttgacgata ttgacgctag ttcccacct agtggttcca 1860
 taccatcaca gacattacca agcacaacag aacaacaggg agccctatct tcagaactgc 1920
 caagcacatc accttcatca gttgctgcca tttcatcgag atcagaataa cacaaccat 1980
 ttactcagtc ccggataacct ccagatttgc ccatgcatcc ggcaccaagg cacataacgg 2040
 aggaagaact ttctgtgtg gaaagtgtt tacatcgctg gaggacagaa atagaaaatg 2100
 acaccagaga tttgcaggaa agcatatcca gaatccatcg aacaattgaa ttaatgtact 2160
 ctgacaaatc tatgatacaa gttccttattt gattcatgc cgttttagt cacgaaggcc 2220
 aagctaattgc tggcactac tggcataat ttttgatca tcgtgaaagg agatggatga 2280
 agtacaatga tattgtgtg aaaaaatcat catggaaaga gctagtggg gactctttt 2340
 gtgggtatag aaatgccagt gcatactgtt taatgtacat aaatgataag gcacagtcc 2400
 taatacaaga ggagtttaat aaagaaactg ggcagcccc tgggttata gaaacattac 2460
 caccggattt gagagattt gttgaggaag acaaccaacg atttggaaaaa gaactagaag 2520
 aatgggatgc acaacttgc cagaagctt tgcaaggaaaa gcttttagcg tctcagaaat 2580
 tgagagagtcc agagacttct gtgacacag cacaaggcgc aggagacccca gaatatctag 2640
 agcagccatc aagaagtgtt ttctcaaagc acttggaaagg agaaaactatt caaataattt 2700
 ccaaggccatc acatgagcat gaaagaaaaa gtcctgaaac agttttcgat tcggcaatta 2760
 agttgaata tgcaagggtt gttagttgg ccaagaaga caccacca gaaaccgatt 2820
 atcgtttaca tcatgttagt gttctacttta tccagaacca ggcacccaaag aaaattattt 2880
 agaaaaacattt actagaacaa tttggagata gaaatttgg gtttggatgaa aggtgtcaca 2940
 acataatgaa agttgtccaa gccaaactgg aaatgataaa acctgaaagaa gtaaaacttgg 3000
 aggaatatga ggagtggcat caggattata gggaaatttgc gggaaacaaact atgtatctca 3060
 taatgggctt agaaaaattt caaagagaaa gttatataaa ttccttgc ttcctcatct 3120
 gtgttcatca gaaataacaaa gaacttgc ttaaggctt atacagagga catgtgaaag 3180
 aatttgcattt acattataga agagaatgtt tgctaaaaattt aaatgagccaa gcccgagaac 3240
 tcttcgaatc tggagaggat cgagaagtaa acaatggttt gattatcatg aatgagttt 3300
 ttgtcccattt tttgcattt ttacttgc tttttttttt agaaaaggat atactagctg 3360
 tagaagatat gagaatcga tgggttccctt accttgc tttttttttt agaaaatggaa ccacacccctcc 3420
 aagaaaaagct gacagattt ttggccaaac tgcttgcattt ttctatggg attaaaagtt 3480
 tccatgagcc accgaaagtta ccttcattt ccacgcattt gttttttttt gttttttttt 3540
 gaatcatgtt gttttttttt cttttttttt ccacgcattt gttttttttt gttttttttt 3600
 gaacacactgt tataaaactt tttttttttt taacccttgc cttttttttt cttttttttt 3660
 ttgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3720
 tatacagact ttatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3780
 ttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3840
 ttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3900
 ccatttcagtc tggatcccttccatgactaca gccatataag tttttttttt tttttttttt tttttttttt 3960
 acataatattt tggtagctt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4020
 aaaaaaaaaaaaaaaa 4028

<210> 16
 <211> 1422
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1856589CB1

<400> 16
 ggcggggca ggcagggtgg gtgcgcaggagg aggctacac tgccttccc ctccgcgtc 60
 ccctcaggcc gaggcggcca ggaccccgga gcgagcgat gggagccccc acctgttaggg 120
 gtcaggat ccccaagcggc ccccccagtcc agggggaaacg cagtgcgecc cgtttcggt 180
 ttacttccctt cggccctgtgg ccagccggat tcaaggataaa ctgggaggattt gccggctcca 240
 gacaggaatg gggccctggca ggtggccctg cagaccacca acagacacat ctgcggaggc 300
 tcccttacgtt ccacacactg ggttataaaagg aggacacaaac caatccagtt tttttttttt 360
 cctgggttc ggaggattttt gaaatgaccc gccattggcc ctgggggtt agcctccgg 420
 tggaaaatga gcacgtgtt ggagggggcc ttatttgcacc cagctgggtt gtactgcgg 480
 cccactgcaccccaaaagactt cttttttttt tttttttttt tttttttttt tttttttttt 540
 ccatgaaactt cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600
 actggggccg ggccttcattt atgggtgc tttttttttt tttttttttt tttttttttt tttttttttt 660
 ctttcatttgc gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
 ggacgcaggatg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780
 ccaactccat gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840
 ggtgtgaccg gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 900
 acatgatctg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 960

tggcttgtga agttgaggggc agatggattc tggctgggggt gttgtcctgg gaaaaggcct 1020
 gcgtgaaggc acagaatcca ggtgtgtaca cccgcgtcac caaatacacc aatggatca 1080
 agaagcaa at gagaatgga gccttctcag gtcctgtgc ctctgcgtc ctctgttcc 1140
 tgtgtggcc gtcgcagccc cagatgggt ctcgcactcc ctaccttcc ctctctgc 1200
 ctgcctctg ctgaatgggg ccagatgggt tgaccaaggt catgtgtcca tcttcaaaaa 1260
 gagtcagggt ggggaagagt aaccctggg agaatgggtc tggcttggc atcccggtga 1320
 ggagaagtgt ggtggatgac taggccttgg gtgagcagga gaaggaaat gtggcctaga 1380
 aggattctgg aatctggac caggagagca gggattaaac at 1422

<210> 17
 <211> 1911
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2617672CB1

<400> 17
 cccacgcgtc cgccggcggt cgcaagacca ggaggcggag ggcgcggggc cagcctggc 60
 cccagccac accttccacca gggcccgagg gccaccatgt ggcgatgtcc actggggcta 120
 ctgcgttgc tgccgctggc tggcaactt gctctgggt cccagcaggg tcgtggggc 180
 cgggagctag caccgggtct gcacccgtcgg ggcacccggg acgcggggagg ccggtaactgc 240
 caggcaggcgg acctgtgtcgg cgcggccgt ggcacgact gtgcctgtcc ctacctggc 300
 gccatctgtt actgtgtaccc tttctgtcaac cgcacgggtc cccactgtc cctgtacttc 360
 tgggacttct gcctcggtt gccacccctt tttccccca tccaaaggat tatgtatgg 420
 ggtcgatctt atccagtcgg gggaaacgtac tgggacaact gtaaccgtt caccgtccag 480
 gagaacaggc agtggcgtg tgaccaagaa ccatgcctgg tggatccaga catgtatcaa 540
 gccatcaacc agggcaacta tggctggcag gctgggaaacc acagcgcctt ctggggcatg 600
 accctggatg agggcattcg ctaccgcctg ggcacccatcc gcccatttc ctcggcatg 660
 aacatgcatg aaatttatac agtgcgtac ccaggggagg tgcttccac agccttcgag 720
 gcctgtgaga agtggcccaat cctgttcat ggccttgc tccaaaggca ctgtgcaggc 780
 tcctggccct totccacagc agtgcgttca tccgtatgtt tctcaatcca ttctctggg 840
 cacatgacgc ctgtcctgtc gcccagaacat cttgtgttcc gtgacacacca ccagcagcag 900
 ggctgcgcg gtgggggtct cgatgggtcc tgggtgtcc tgcgtgcgg aggggtgg 960
 tctgaccact gtcacccctt ctggggcggt gaacgagacg aggctggccc tgcgc 1020
 tgtatgtatgc acagccgagc catgggtcgg ggcacccatgc aggccactgc ccactgc 1080
 aacagctatg ttaataacaa tgacatctc cagggtactc ctgttaccc cctcggttcc 1140
 aacgacaagg agatcatgaa ggagctgtat gagaatggc ctgttcaaggc ctgtcatgg 1200
 gtgcgttgcg acttccctt atacaaggga ggcacccatc ggcacacgc agtgcgttcc 1260
 gggaggccag agagataccg ccggcatggg accccactcag tcaagatcac agatgggg 1320
 gaggagacgc tgccagatgg aaggacgcctt aaatactggg ctgcggccaa ctctctggg 1380
 ccagcctggg gcgagagggg ccactccgc atctgtgcgc ggcgtcaatga gtgcacatc 1440
 gagagcttcg tgctgggggtct ctggggccgc gtgggcattgg aggacatggg tcatactga 1500
 ggctgcgggc accacgcggg gtccggctt ggcacccatgc taaggggccgg cgaaagaggc 1560
 cccaaatgggg cgggtaccc acccttgcggc gacagaccc ggggcgcagg cgggcgc 1620
 ggcgttacatc cggcgcgggg ttccgttgc gacgcgttcc ggcctgggaggc cggggcagg 1680
 cgagactggc ggagcccgac acctcccaat gggggacgggg cagggccctgg ccttggg 1740
 gcacagctgc agatcccagg cctctggcgc cccactcaa gactacccaa gccaggacac 1800
 ctcaagtctc cagccccact accccacccc accccctgtat tcttatttcc tcaatattta 1860
 ttttttttta cactgtttta aaataaaaacc aaagtattga taaaaaaaaa a 1911

<210> 18
 <211> 854
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2769104CB1

<400> 18
 caccttttgt tcccttatectt gggccagttc tctcgatgtt cccagatgtc cagttccaga 60
 tgcctggacc cagagtgtgg gggaaatatc tctggagaag ccctcaactcc aaaggctgtc 120
 caggcgcataat gtgggtggctg cttctgtgg gatgttccca ggcttgccttca accccggggct 180
 ccgttctctt ggcccaagag ctaccggcgc agtgcacatc ccccggttgc ccagagccgt 240
 atggcaagg ccaagagacg acacggaca tcaaggctcc agagggtttt gctgtgaggc 300

tcgtttcca ggacttcgac ctggagccgt cccaggactg tgcaggggac tctgtcacaa 360
 ttcattcgt cgggtcgat ccaagccagt tctgtgtca gcaaggctcc cctctggca 420
 ggcggccctgg tcagaggag tttgtatcct cagggaggag tttgcggctg accttccgca 480
 cacagcccttc ctggagaaac aagactgccc acctccacaa gggcttccctg gccctctacc 540
 aaaccgtggg ttagtgtccc tcctgggggt gcagggaggg agcctctgtt cccagccatg 600
 accctggtat ttcaagcct taagtggaaag ttgagtgac agctgaggct ggggactcag 660
 ggacacctgg gctggatccc agccctgccc ctgctggcaa gcaaccctat taagagacag 720
 ccgtagctga gccccccagcg gttgttcca tgcagattt cagggccagt gttgcagat 780
 catctcatc ttaaagagat gccaaaatc cagattttt agtaaaaat taaattttca 840
 aaaaaaaaaa aaaa 854

<210> 19
 <211> 1386
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4802789CB1

<400> 19
 gacgctgcgg cccggcccg cgggtaaata acagatgcgg gtgaaagatc caactaaagc 60
 tttacactgag aaagccaaaaa gaagtaaaag gcctactgtca cctcatgtatc aagactcttc 120
 agatgatatt gctgttagtt taactgtcca acatgtttagt catgtatca gctgtatca 180
 tgtaaagaga gcaatagctg agaatotgtg gtcagtttc tcagaatgtt taaaagaaaag 240
 aagattctat gatgggcagc tagtacttac ttctgtatatt tggttgtgcc tcaagttgtgg 300
 cttcaggga tggtaaaaaa actcagaaag ccaacattca ttgaagcaact ttaagagttc 360
 cagaacagag cccattgtt ttataattaa tctgagcaca tggattatat ggtgttatga 420
 atgtgatgaa aaattatcaa cgcattgtttaa taagaaggat ttggctcaga tagttgat 480
 tctcagaaaa catgttctca aaacacaaac aagtgcattt tctagaatca tggaaactttg 540
 tgaagaaaaa tggtaaacag atgaaataca gaagggagga aatgcagaa atttatctgt 600
 aagaggaatt acaaattttag gaaatacttg ctttttaat gcaatgtc agaacttggc 660
 acagacttat actcttactg atctgtatca tggatgttcaaa gaaagtagta caaaactcaa 720
 gattttcctt tcctcagact ctcagctggc cccattgggt gttggactttt caaggcctgg 780
 accactgacc tcagccttgc ttctgtttt tcacagcatg aaggagactg aaaaaggacc 840
 actttctcctt aaagttcttt ttaatcagct ttgtcagaag tgggtgcac tacatttaat 900
 ataaataattt atgagttaca aaataactaat gtattcatca tttaacatga atagtcgttt 960
 ttactgttac ttgcctctt ttgccttgac tatgaagaga actaaaattt gttacagctc 1020
 tatgtttat gaaaattata ttcagtcctt cagaagaagc agcttacatc catatataag 1080
 gaaatggaga cacagaaaattt aatggctca cctagtctga gtggaaagct gagaatcaa 1140
 tggagatctg tcctgacttg gatgcctatg ttgtatacc ataaagttagt aaaaaccatag 1200
 agttgtaaaaa tctagaaaatg accgttaatg aacatctaat ctatgtttt tattttaaaa 1260
 gatgagctgt gaggcaataa gaggtaatgtt gatgttca aggtattaca gtatgtttaa 1320
 aaaccaaaatc ttatgtgcc tggaaataaa cacataaagg atctgacttg aaaaaaaaaa 1380
 aaaaaaaaaa 1386

<210> 20
 <211> 3323
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 60116897CB1

<400> 20
 caaatctgca gcagcatgat ttaagattaa attcatgtat tggaaatattt gttcagaccc 60
 catgtgacat aactggagcc agtgcagtgc catgaagaa tacgagatta gcctggatatt 120
 taacttgtct tctagagaat agatttcatg ttccattctt ctgcaatgtt taattcacac 180
 agaaaaccaa tggtaatcat tcacagagga ttttactgtct taacagccat ctggcccaa 240
 atatgcattt gttctcagg ttcagtgcca tctgttattt acttcaactgaa ggtatcctggg 300
 gctttccag tagccactaa tggggaaacga ttcccttggc aggagctaa gctcccccagt 360
 gtggatcttccatc tgaccctttt gtccacccca atctcacctc tctggactttt 420
 gttgcattgtg agaagatgtt agtctggtc agcaatgtca cccagttat catcttgcac 480
 agcaagatc ttgaaatcac gaatgcaccc ttctcgtca aggaagatc aagatacatg 540
 aaaccaggaa aagaactgaa agtttgagttt tactatgtgg ttcactgtctt atgaacaaat tgcactgtcg 600
 gttccagaga aacttacgccc tcacctgaaa tactatgtgg ctatggactt ccaagccaaag 660

<210> 21

<211> 2123

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1866356CB1

<400> 21

tgacaatcca	agatggcgg	gcccggcgag	gcggaggagg	aggcgacagt	ttacctggta	60
gtgagcgtta	tccccctccgt	gttgcgtctcg	gcgcatttac	ggracttatt	tagccagttc	120
cgagaagagc	ggggcgggtgg	cttcctctgt	ttccactacc	ggcatcgccc	tgagggggcc	180
cctccgcagg	ccgtctctaa	ctctgcoccta	attctaccac	acccagccgc	tgagggccag	240
cttctcttc	agacttcggc	caccatgttc	cggctctct	caactcgaga	ctctactcca	300
atccagaccc	gcacactgctg	ctgcgtcatc	tcggttaaggg	ggttggctca	agctcagagg	360
cttattcgca	tgtactcggg	ccgcgggtgg	ctggattctc	acgggacttg	gctaccgggt	420
cgctgtctca	tccgcagact	tcggctaccc	acggaggcat	caggtctggg	ctccttcccc	480
ttcaagaccc	ggaaggaaact	gcagagttgg	aaggcagaga	atgaagccct	caccctggct	540
gacctgaagc	aactgcggga	gctgaaccca	ccagtgtctga	tgcccagagg	aatgtgggg	600
actccccctgc	gggttttttt	ggagttgatc	cgggcctgccc	gcctacccccc	tcggatcatc	660

acccagctgc agctccagg tcccaagaca gtttcctccc ggcgctacgg caatgtgcct 720
 tttgagtagt aggactcaga gactgtggag caggaagagc ttgtgtatac agcagagggt 780
 gaagaaatac cccaaaggaa ctacctggca gatataccag ccagccctg tgagagac 840
 gaggaagaag tggggaaagga agaggaagaa gagtctact cagatgagga ctagac 900
 ggtgaggaat gggaaacggca tgaagcgctg catgaggac tgaccggca ggacggacc 960
 actgagcagc tctttgagga ggagatttag ctcaagtggg agaagggtgg ctctggcctg 1020
 gtgtttata ctgatgcca gttctggcag gaggagaag gagatttga tgaacagaca 1080
 gccgatgact gggatgtgga catgagtgt tactatgaca gagatgggtt agacaaggat 1140
 gcccggact ctgtccaaat gctgtggaa cagagactcc gagatggaca ggaagatggc 1200
 tctgtatcg aacgcaggat gggcacctt gacgcacca ccaaggcat tggcgaaag 1260
 gtgtggagc ggcagggtg ggcgtgaggc caggcctgg gctgcagggtg ctcaagggtg 1320
 cctgaggccc tggatagtg tggccaaac cccagatgca agcgtggatt ggggttccat 1380
 ggagagaagc tacagccatt tggcaactg aagaggcccc gtagaaatgg ctggggctc 1440
 atctccacca tctatgtatg gcctctaccc caagaccaga cggagtca gctccggcgc 1500
 cagccaccca ccagcatgaa gtttccgaca gacatggcct ttgtggggg ttccagttgt 1560
 gcttcagaca gccccttatt gcctgactga ccgggttggg ggcttcattt catagctaca 1620
 tgatgaaaac cctctggcct ggcctcatct accactgaaag cagaaaggag tctggggagca 1680
 gcagtcctcg tggctggttc agggtttt gttccgagcc tgcctggctg ccgggttctat 1740
 acctcagggg catttttaca aaaagcccccc tccctggccc tcccttggg tattagggtt 1800
 aacgaccgct tgcgtttgtt ctctaaacctt aatctctggg ctggccctt gcctctgca 1860
 gaaccttggaa aagctgggtt gagtgggtt atcagcacag cttcccttgg ggactctgaa 1920
 ggtgtcccca cgaaggccag aaagggggaa agggacctgg gcgaggagag gatttgggtt 1980
 gcttggaaaga gcccggctt ggtggccctt ccacggcctc tacccttact gggtggact 2040
 gccagcggag agtccgggg aggtggctt ggtgtgcac gtcacggaa aataaagacg 2100
 tttactactt gaaaaaaaaaaa aaa 2123

<210> 22
 <211> 2893
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1872095CB1

<400> 22
 atgcatcatt tgaaccttct gtagcattgg caagccttgc gcagcatatt ccttcata 60
 tgattacagt tctcatcagg agcctacta cggatccaa tgaaaaatg gcaagttatg 120
 cccaaaggcc ttgcagaatg attgacttgc tattctggc attggctcag catgtggata 180
 catggtaat tgcactctg aaaggactgg cagctgtcca gaagtttact attttgatag 240
 atgttacttt gctgaaaata gaactgggtt ttaatcgact ttggtttctt ctgtgagac 300
 ctgggtctct tgcagttctt tctcacatgc tgcttagctt tcagcattct ccagaggcgt 360
 tccatggat tggcattcat gtggtaatt tggttcatcc ttccaaaaat gatggctcgc 420
 cttcaagttac agccttcttta gtacaattaa cagaattgtt acactgtatg atgtatcatt 480
 attctggatt tccagatctc tatgaaccta ttctggaggc aataaaggat ttccataa 540
 ccagtgaaaga gaagatggaa ttaattctca tcaaaatgtc ctggacttct caatccaaatt 600
 ctggcgctc ttgtttgtt agacttctg gaaaatctgaa aactggggaa actggctta 660
 ttaacctagg aaatacatgtt tataatggaa gtgttataca agccttgggtt atggccacag 720
 atttcaggag acaagttatc tcttttataatgggtt caattcatta atggaaaaat 780
 tacagcatct ttttgcctt ctggccata cacagaggga agcatacgca ctcggatata 840
 tctttgaggc ttccagaccc ccatggttt ctcccaagatc acagcaagac ttttctgaaat 900
 acctcagatt ttccttgcac aggctccatg aagaagaaaa gatcttggaa gtcaggccct 960
 cacacaagcc ttctgaaatt ctggaaatgc gtggaaactt tttacaggaa gtagctgtt 1020
 aagcagcagt actaaacagag acccctcgta caagtgcgg tgagaagact ttaatagaaa 1080
 aaatgttgg agggaaaacta cgaactcaca tacgttggta gaactgcagg agtacctcact 1140
 aaaaatggtaa acgccttataca gatcttcgc ttgccttttgc ttcttcctct tctttggaaa 1200
 acatgtctgt ccaagatcaca gcatcatcac ccagttataca agatgggtt ctaatgcac 1260
 cctctgttacc cggcccttca gaagaaccag tagttataaa tccaacaca gtcgcctca 1320
 tctgtgactc acttgcattt gaaaaatcca taggcgttcc tccataatgat ttttactgtt 1380
 ctgaaaacac ttctgtccctt aacgaatcta acaagattct tggtaataaa gatgtaccc 1440
 agaaaccagg aggtgaaacc acacccctcag taactgcatt actaaattat ttttggctc 1500
 cagagattct tactgggtat aaccaatattt attgtggaaa ctgtgcctct ctgcggaaatg 1560
 ctgaaaaac tatgcaaatc acggaggaaac ctgaaatccat tattttactt ctccctgagat 1620
 tttcatatgtt tcaagatgtt catgtggaaa gggaaaattttt agacaatgtt tcaactgcac 1680
 tggtttggg gttggcaggat aaaagaattt ctttttctt ttcattgtca gaaagttgg 1740
 ctgtatgtt tgacttcaactt gatcttagt agaacccttgc taaaaattt aaggccctc 1800
 ggactgtatgtt agtcccttgc aaaaaatttgg tggccatctt attaagttcc gttgtgggtt 1860

actctggat	atccctctgaa	agtgggcatt	actattctta	tgccagaaaat	atcacaagta	1920
cagactttc	atatcagatg	taccaccagt	ctgaggctct	ggcattagca	tcctccccaga	1980
gtcatttact	agggagagat	agtcccagtg	cagttttga	acaggatttg	aaaaataagg	2040
aaatgtcaaa	agaatggttt	ttattnaatg	acagtagagt	gacattact	tcatttcagt	2100
cagtccgaaa	aattacgagc	aggtttccaa	aggcacacgc	ttatgtgcct	ttgtataaaa	2160
aacagcatag	tactaatggt	ttaagtggta	ataacccaaac	cagtggactc	tggataaaatg	2220
gagaccacc	tctacagaaaa	gaacttatgg	atgcataaa	aaaagacaat	aaactatatt	2280
tacaggaaca	agagttgaat	gctcgagccc	gggcctcca	agctgcatct	gcttcatgtt	2340
catttcggcc	caatggattt	gatgacaacg	accacccagg	aagctgtgga	ccaaactggtg	2400
gaggggggtgg	aggaggattt	aatacagttt	gcagactcgt	attttgcattc	tgagagagtc	2460
caaaatgcac	tggtcacgaa	acgtctaata	ctatgactgt	taaaaatgtca	gactataaca	2520
aatatctatc	tttttatttt	cattagaccc	ttatacttca	agagaacaca	ctcagtgctt	2580
gttttttattt	tcttgacaca	tttattaaaca	aaatgcatca	tggaaaaaaa	aatctacctc	2640
ttaaaaatcc	atttgtttt	atggtagac	atgcgttgcacc	aaaaatgttc	agaagaaaat	2700
atgtacttgg	tccttaatta	agctgcgtt	aatttggtag	aagcattttaa	atggtctatc	2760
ttcagttta	ctgaacaaaa	aatgtattt	attttagattt	ctttaaaaaa	gaattgtatgc	2820
tagaggtaaa	aaaaaaatact	tgtttttaaa	aaatccttta	cgtcttgcgt	aattaccccg	2880
attattaaat	tca					2893

<210> 23
<211> 4170
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 2278688CB1

400 > 23	gctccccccgg	tcgtctctct	ccggcggtcg	cccgcgctcg	gtggatgtgg	cttgcagctg	60
ccgccccctc	cctcgctcgc	cgcctgtct	tcctcgcccc	tccgcctcct	cccccctcc	120	
tttcgtcttt	cagccgctcc	tctcgccgccc	gcctccacag	cctggccctc	gcccgcgtgc	180	
cggagaagag	gcccgtcgag	cggctgcctg	ccgtatgtctc	ccccatcaac	tgcagccctt	240	
gcctcaagcc	cgatgtctg	gacttcacct	tcgagggcaa	gctggaggcc	gcccgcctagg	300	
ttggccggc	gactaatcag	attgtatgat	attgtgctga	tattgatatt	attacagctt	360	
catatgcacc	agaaggagat	gaagaaatac	atgcacagg	atthaactat	cagaatgaag	420	
atgaaaaagt	cacccgtct	ttccctagta	ctctgcaaac	aggtacggga	accttaaaga	480	
tagattttgt	tggagagctg	aatgacaaaa	tgaaaaggttt	ctatagaagt	aaatatacta	540	
cccttcgtgg	agagggtgcgc	tatgctgtg	taacacagt	tgaggctact	gatgcccggaa	600	
gggttttcc	ttgtcgggat	gaggctgta	tcaaaacac	ttttgatata	tcattgttgc	660	
ttccctaaaca	cagagtagct	ttatcaaaca	tgaatgtaat	tgaccggaaa	ccataccctg	720	
atgatgaaaa	tttagtggaa	tgtaagttt	cccgacacac	ttgttgcgtt	acatatctgg	780	
ttggcatttgt	tgtgggtgaa	tatgacttt	tagaaacaag	gtcaaaaagat	ggtgtgtgtg	840	
tccgtgtta	cactccgtt	ggcaaaagcag	aacaaggaaa	atttgcgtt	gaggttgctg	900	
ctaaaacctt	gccttttat	aaggactact	tcaatgttcc	ttatcctcta	ctaaaatttg	960	
atctcatgtc	tattgcagac	tttgcagctg	gtgcacatgg	gaactggggc	cttgttactt	1020	
ataggggagac	tgccattgtt	attgtatccaa	aaaattccctg	ttcttcatcc	cgccactggg	1080	
ttggctctgtt	tgtgggacat	gaactcgccc	atcaatgttt	tggaaatctt	gttactatgg	1140	
aatgggtggac	tcatcttttg	ttaaatgaag	gtttgcata	ctggattgaa	tatctgtgtg	1200	
tagaccactg	cttcccccagag	tatgatattt	ggactcagtt	tttttctgt	gattacaccc	1260	
gtgcccagga	gcttgcaccc	ttagataaca	gccatcctat	tgaagtca	gtgggcccatt	1320	
catctgaggt	tgtatgagata	tttgatgcta	tatcatatag	caaagggtca	tctgtcatcc	1380	
gaatgtctca	tgactacatt	ggggataagg	actttaagaa	aggaatgaac	atgtattttaa	1440	
ccaaatgttca	acaaaaagat	gctgccacag	aggatctctg	ggaaaagttt	aaaaatgcta	1500	
gtggtaaacc	tatgcagct	gtgtatgata	cctgaccaa	acaaatggga	tttcccctca	1560	
tttatgttga	agctgaacag	gttagaagat	acagattatt	gagggtgtcc	aaaaagaagt	1620	
tctgtgtctt	ttggtcata	gttggatgaa	attgtcccc	ttggatggc	cctatcacaa	1680	
tctctactag	tgaagacccc	aaccaggcca	aactaaaaat	tctaattggc	aagccagaga	1740	
tgaatgttgt	tttgaaaaat	gtcaaaaccag	accaatgggt	gaagttaaac	tttaggaacag	1800	
ttgggtttta	tcggacccag	tacagctcg	ccatgctgga	aagtttatta	ccaggcattc	1860	
gtgaccttcc	tctgccccct	gtggatcgac	ttggattaca	gaatgacctc	ttctcccttgg	1920	
ctcgagctgg	aatcattagc	actgttagagg	ttctaaaagt	catggaggt	tttgcataat	1980	
agcccaattt	tactgtatgg	agcgcaccc	gctgtaccc	ggggattctc	tcaactctct	2040	
tgtcccccac	agactttat	gaggaaatcc	aggagttgt	gaaagatgtc	ttttcaccta	2100	
tagggggagag	actgggctgg	gaccggaaat	ctggagaagg	tcatctgtat	gcactctgt	2160	
ggggctttgt	tctggaaaaa	ctagggaaaag	caggacataa	ggcaacgtt	gaagaagccc	2220	
gtcgctcgat	taaggaccac	gtggaaaggaa	aacagattct	ctccgctgat	ctgaggagtc	2280	

ctgtcttatct	gactgttttg	aaggcatgggt	atggcactac	tttagatatt	atgtttaaaac	2340
ttcataaaca	agcagatatg	caagaagaga	aaaaccgaat	cggaaagatc	cttggcgcta	2400
ctctttgccc	tgaccctgatt	caaaaagtcc	tcacggttgc	actttcagaa	gaggtagcgtc	2460
cacaggacac	tgtatcggtt	attgggtggag	tagctggagg	cagcaagcat	ggttaggaaag	2520
ctgcttggaa	atttcataaag	gacaactggg	aagaacttta	taaccgcatac	caggaggat	2580
tcttaatatac	cagactaata	aagctatcatg	ttgagggatt	tgcagtttat	aaaatggctg	2640
gagagggttaa	ggctttcttc	gagactcacc	cgactccttc	agctgagcgt	accatccagc	2700
agtgttgcga	aaataattctg	ctgaatgtcg	cctggctaaa	gcgagatgct	gagagcatcc	2760
accagtacct	ccttcagcgg	aaggcctcac	caccacagt	gtgaatcccg	aggtgccgccc	2820
attggcggtt	ctgtcgttc	gctgcaggg	taaggtggag	ctaccgaaca	gctgattcat	2880
atgccaagaa	tttggagtct	tctttcaaaac	cagtgggggt	tggacaatga	atgttagttaa	2940
ctggttcccg	ctcacactcc	agaattaaat	tctattggaa	aaggaaaatc	agcaattcag	3000
caaaaaataaa	ataaaaaataa	aaaatgtaaa	tatgataga	ataaaaataga	gcataacgaa	3060
actgtgaac	ttttgtgaagc	cttgcagt	gtttaaaggta	tttaaactc	tacttttaat	3120
gacagatgtt	ctgtttttat	aacctaccaa	aaggaaacta	gaggcttctt	ggtgaagagc	3180
atttttgtga	agtgggttct	gcaaggagcc	tataaagcc	agggtgggtt	ccatttctgg	3240
gaatgggttaa	acacaaaagg	ctgatagctg	gtatcacata	gttggagtc	gtgcataatt	3300
ccaagtggct	tttttttttt	ttggcacggg	gactgatcag	gaagatata	tcctgcataa	3360
ctcaatctga	accaaggatt	gtagtttagt	tttcctccct	gcctccctt	ctgtgtgacc	3420
gacccttgg	ccaaaaaaa	aaacaaaaaaag	caaaaaacaa	aaacccatccc	tgttctggtt	3480
ttttccctcc	ctttagttcc	accccccaccc	cccatccct	ggttccttc	tttagagatga	3540
agaaaataata	agggaaacatc	tttcataggc	acattaaata	agagaaaactg	atatacatta	3600
ttttttctt	tttaaagatg	actttaaaga	acccctaaaat	ttatataatgt	gagacaatag	3660
aaataaaaaaag	atcttcagcc	aggcctttct	gaaggaggtt	ttctgtctaa	aatggcttta	3720
gttgtctgaa	aagccagtc	ttgaacctct	tcacaacagt	atcaacactg	gcttcctcccg	3780
gttcattttt	tgcggtcgag	aagtcaatgg	taactgctgc	agggttttaat	acatttagtgg	3840
taactgggtt	aaaaaaacaaa	gactgtaaac	ctgtgtgtc	cactgtttc	ttcaacacgt	3900
tatcctacta	ataagcctca	ccttatttaat	ccaaatggat	ttaaatctaa	atctcattcc	3960
cttcttcctt	ccctacccctt	ttttttctt	ttttttttttt	aaatattttt	tgtttattaa	4020
agaaaattcat	atttgggtgt	gctttaacgg	atttcagaag	gtcatcagat	tgtgagactg	4080
cttccttggaa	acatttttgt	gctattgtt	aaaaaaaata	ataaaaaaac	agttggcggtt	4140
aataaaaaatg	tcaatgtgaa	aaaaaaaaaa	aaaaaaaaaa	ataaaaaaac	ataaaaaaac	4170

<210> 24
<211> 767
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 4043361CB1

```

<400> 24
ccgagaggt gcagcggcac agctgtcgcg ccagtcgcaa cagaagcagg tccgagggcac 60
agcccgatcc cgccatggag cagccgagga aggccgtgggt agtgacggga ttggccctt 120
ttggggaaaca caccgtgaac gccatgtggta ttgcattca ggagctagaaa aagctaggcc 180
ttggcgacag cgtgacactg catgtgtacg agatccgggt tgagtaccaa acagtccaga 240
gactcatccc cgccctgtgg gagaagcaca gtccacagct ggtgtgtcat gtgggggtgt 300
caggcatggc gaccacagtc acactggaga aatgtggaca caacaagggc tacaaggggc 360
tggacaactg ccgttttgc cccggctccc agtgtgcgt ggaggacggg cctgaaagca 420
ttgactccat catgacatg gatgtgtgt gcaagcgagt caccacgtt ggcctggatg 480
tgtcggtgac catctcgcag gatgcccggca gggaaaaacc ctccctgc aaagggtact 540
gtgttttctg ccgccgaagg agggcccggt ccctccaggc tcagtgtggc ttctccctga 600
cccccgccctt agaaactttt ccagtgcctt ttctgaaact cttgtgtccccc gggccccc 660
ggcggagaag gatagtccggg attctgcctt gggctggct ctaggagacc ccaaatttga 720
caccacagaa agcaaataaa acacttgaaa tacgcaaaaa aaaaaaaaaa 767

```

<210> 25
<211> 1538
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<223> Incyte ID No: 3937958CB1

<400> 25

ggtgagtggg aggcatgggg tggatgagaa gcctaggcag aggctttcc tgcacatcc 60
 ctcagttcc ttattcacag atgcggccct ccctgtctac ctgtatgaat ttgagcacca 120
 cgctctgga ataatcgta aaccccgac tgatggggca gaccatgggg atgagatgta 180
 cttctcttt gggggccct tcgccccagg tgcaaagggtc ccacctgata ccccaactgg 240
 gtgtccagtc tcccacctt ggatgcagac ccacccctcc attggctgc cacagggagc 300
 tcaccagttc ctaatctgtt atgctctccc aaatgaaagt cttctgtcc ggaagcagca 360
 gaagcagcag gagtaggggtg ggaggtcagt gtcccctgct ctgtccgaaa tcccacatcc 420
 cattctgccc ccaggccctt ccattggtaa ggagaaggca cttagccctc agatgtgaa 480
 atactggcc aactttgccc gcacaggaaa ccccaatgat gggaaatctgc cctgctggcc 540
 acgctacaac aaggatgaaa agtacccgtca gctggatttt accacaagag tgggcatgaa 600
 gctcaaggag aagaagatgg ctttttggat gagttctgtac cagtctcaaa gacctgagaa 660
 gcagaggcaa ttctaagggtt ggctatgcag gaaggagcca aagaggggtt tggcccccacc 720
 atccaggccc tggggagact agccatggac atacctgggg acaagagttc taccacccccc 780
 agtttagaac tgcaggagct ccctgtgcc tccaggccaa agctagagct ttgcctgtt 840
 gtgtgggacc tgcactgccc ttccacggct gacatcccat gatgcccctc tacttcaactg 900
 ttgacatcca gtttagggccag gcccgtcaa caccacactg tgctcagctc tccagccctca 960
 ggacaacccctt tttttccctt ttcttcaaat cttcccccacc ttcaatgtct ctttgtgact 1020
 ccttctttagt ggagggtcgc ccagactgcc actgccccctg tcactgcacc cagcttggca 1080
 tttaccatcc atccctgtctca accttgcgtcc tgcgtgttca cattggctgt gaggcctagg 1140
 gcaggtgtg acatggagca aactttgggt agtttggat cttctctccc acccacactt 1200
 atctccccc gggccactcc aaagtctata cacaggggtt gtcttctcaa taaagaagtg 1260
 ttgattagac ctgaatttct ccacctataa aatgggtgtg tgaagtgaat gatgtctcaa 1320
 tttagggccctt gagagaaaagg aagtattgtc gcctgttccct tagtgggtgt tgccctggatg 1380
 ctacactcg tcaaagggtg ctactgcaaa gttgcctggg gtacaaaaca ctgcctttg 1440
 gcccctcatg gtctcaagt caccctcag gacagccaca ccccacgctc acttgtccat 1500
 cagtttaggt tttagtgcctt catctagatt cctctggc 1538

<210> 26
 <211> 1497
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7257324CB1

<400> 26
 ggccttactc ttccaagagg ccatggaagt ataaataata aagcaagaaa ggcagatgca 60
 tttggctggc tcagtgact tctgaatgtt ctgtgagttt gagaccccttcc cttccaaaag 120
 atccgggtct ttttgtctat tccacacgaa gcttgcttca gatcgaggga ggtatgtagca 180
 ctgtccacag gtctactact caacaggata ttcttcaagg aaaatgaaacc cccacactagg 240
 cctggccatt tttctggctg ttctcttcac ggtgaaagggtt cttctaaagc cgagcttctc 300
 accaaggaat tataaaagctt tgagcgaggta ccaaggatgg aagcaaaaggaa tggcagccaa 360
 ggagcttgca aggcaagaaca tggactttagg tttaagctg ctcaagaagc tggcctttta 420
 caaccctggc aggaacatctt tcctatcccc cttgagcatc tctacagctt tctccatgt 480
 gtgcctgggt gcccaggaca gcaccctggc cgagatcaag caggggttca acttcagaaa 540
 gatggcagaa aaaaatcttc atgagggtttt cattacatc atccacgagc tgaccaggaa 600
 gaccaggagc ctcaaaactgta gcattttggaa caccgtgttc attgaccaga ggctgcagcc 660
 acacgtaag tttttggaaat atgcaagaa cttttacagt gcccggaaacc tctttaccaa 720
 ctttcagaat ttggaaaatgg ctcagaagca gatcaatgac tttatcagtc aaaaaaccctt 780
 tggggaaaatt aacaacctga tcgagaatat agaccccgcc actgtgatgc ttcttgcaaa 840
 ttatatttttcc ttgcagccca ggtggaaaca ttagtttgat ccaaatgtaa ctaaagagga 900
 agatttttcc ttggagaaaaa acagttcactt caagggtcccc atgatgttcc gtgtggcat 960
 ataccaagttt ggctatgtac ataaatcttc ttgcaccatc ctggaaataac cttaccagaaa 1020
 aaatatcaca gccatcttca tccttcttca tgagggcaag ctgaaagact tggagaagg 1080
 attgcagggtg gacacttttcc ccaagatggaa aacattactg tcacgcaggg tgcgttagacgt 1140
 gtctgtaccc agactccaca tgacgggcccc ctgcacccctt aagaagactc tctccatcat 1200
 aggtgtctcc aaaatcttgc aggaacatgg tgatctcacc aagatgcggcc ctcatcgcc 1260
 cctggaaaatgg ggcggggctg tgcacaaggc tgagctgaaatgatgaga ggggtacgg 1320
 agggcccgctt ggcaccggag cacagactctt gcccacggag acaccactg tcgtcaagat 1380
 agacaaaaccc tatctgtcc tgatttacag cgagaaaata cttccgtgc ttctctggg 1440
 aaagattgtt aacccttattt gaaaataaaag gagaattccctt gttggccaca aaaaaaaa 1497

<210> 27
 <211> 1194
 <212> DNA
 <213> Homo sapiens

```

<220>
<221> misc_feature
<223> Incyte ID No: 7472038CB1

<400> 27
atgccccggg ccattagtc cctgatgagg tttcaacatc cggtcagttt caagctgcag 60
ctgtaccgcg ttccctgcg ccgcgtcccc tccgcggcgc atcgcttcga gaagttggc 120
atccggatgg accggctgcg tttaaagtac gccgaggagg tcagccatt ccgtggcgag 180
tggaaactcgg cggtgaagag cacaccactg agcaattacc tagacgcccc gtactttggc 240
cccatcacca ttggtaacgc gccgcagaca ttcaaggtga tattcgatac gggttctcg 300
aatctctggg tgccatccgc cacgtgtgcg tccacaatgg tggcctgtcg tgcgcacaat 360
cgctactttg ccaaggcgtc gaccgtcac caggtgaggg gagaccactt tgccatccac 420
tatggcagcg gcagttctgtc cggcttcctt tccaccgaca ccgttcgggt ggctggctca 480
gagattcggg atcagacacctt cggcgaggcc accgaaatgc cgggttcccat cttctggca 540
gcaaaaatcg acggcatctt tggattggcc tatecgacg tctctatgcg ggcgcatacg 600
ccaccattct atgcgtatgtt ggagcaaggaa cttctaaca aaccatattt cagtgttac 660
cttagcagaa atggcgaaaa ggtatgtgga gccatctt tttggcggatc caatccgat 720
tactacaccc gcaactttac ttatgtccag gtgagccatc gtgcctattt gcaggtgaaa 780
atggattccag cagttatccg gaatctcgag ctatgtcagc agggatgtga agtgattatc 840
gacacgggca cctcttctt ggcattgccc tacgaccagg ctatacttat caatgaatcc 900
attgggggaa ctcccccttc ctttggacag tttctatgtt cgtgcgcacag cgtaccagac 960
ctgccccaaa tcacccttac cttgggtggg cgtagatttt tcctggagtc tcacgagtat 1020
gtctttcggg atatcttacca ggatcgaaagg atctgtctt cggcggttcat tgccgtggac 1080
ctgccccatcgc ccagtgacc gctctggatt ctggggatg tgtttttggg caaatactat 1140
actgagttcg acatggagag gcatcgattt ggattcgcggc atgcccaggag ttga 1194

<210> 28
<211> 438
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472041CB1

<400> 28
atggggatcg gatgtggag aaacccctg ctgctgtga ttggccctgggt cctgtcagcc 60
aagctgggtc actttccaaag gtgggagggc ttccagcaga agctcatgag caagaagaac 120
atgaattcaa cactcaactt cttcatccaa tcctacaaca atggccgacaa cgacacccatc 180
ttatatacgat tccagagggt aattcgaaat cagatgcgc tgacgcacgg agtggatgt 240
atagtcactt tcaagatggg tggacccaaa tgcaagagga atgacacgcg caattttcc 300
tgccccctgc aaagcaagaa gttggaaaat agtttaattt gcgagtctt gatatacacc 360
atgcccctggaa taaaacttcc ctagtctgg aacaattctt gtctggaggc cgagcatgtg 420
ggcagaaacc tcagatgtga 438

```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/046443 A3

(51) International Patent Classification⁷: C12N 15/57.
9/48, A01K 67/00, C07K 16/40, C12Q 1/68, A61K 38/48,
G01N 33/50

826 Lois Avenue, Sunnyvale, CA 94087 (US). NGUYEN,
Danniel, B. [US/US]; 1403 Ridgewood Drive, San Jose,
CA 95118 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick
Court, San Jose, CA 95118 (US). YAO, Monique, G.
[US/US]; 111 Frederick Court, Mountain View, CA 94043
(US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara,
CA 95054 (US).

(21) International Application Number: PCT/US00/34811

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(22) International Filing Date:
19 December 2000 (19.12.2000)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

Published:

— with international search report

(30) Priority Data:
60/172,055 23 December 1999 (23.12.1999) US
60/177,334 21 January 2000 (21.01.2000) US
60/178,884 28 January 2000 (28.01.2000) US
60/179,903 2 February 2000 (02.02.2000) US

(88) Date of publication of the international search report:
11 July 2002

(71) Applicant (for all designated States except US): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (for US only): YANG, Jun-
ming [CN/US]; 7125 Bark Lane, San Jose, CA 95129
(US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago
Road, San Leandro, CA 94577 (US). BURFORD, Neil
[GB/US]; 105 Wildwood Circle, Durham, CT 06422
(US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle
Lane, Brisbane, CA 94005 (US). LU, Dyung, Aina, M.
[US/US]; 233 Coy Drive, San Jose, CA 95123 (US).
REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue,
#3, Sunnyvale, CA 94086 (US). YUE, Henry [US/US];

WO 01/046443 A3

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

INTERNATIONAL SEARCH REPORT

International Application No	
PCT/US 00/34811	

A. CLASSIFICATION OF SUBJECT MATTER					
IPC 7	C12N15/57	C12N9/48	A01K67/00	C07K16/40	C12Q1/68
	A61K38/48	G01N33/50			

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N A01K C07K C12Q A61K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category ^a	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>VALERO REBECA ET AL: "USP25, a novel gene encoding a deubiquitinating enzyme, is located in the gene-poor region 21q11.2" GENOMICS, ACADEMIC PRESS, SAN DIEGO, US, vol. 62, no. 3, 15 December 1999 (1999-12-15), pages 395-405, XP002153277 ISSN: 0888-7543 the whole document</p> <p>---</p> <p>-/-</p>	1-7, 11-15

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

*** Special categories of cited documents :**

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

4 September 2001

Date of mailing of the international search report

09.01.2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

ESPEN, J

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/34811

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	GROET J ET AL: "NARROWING OF THE REGIONS OF ALLELIC LOSS IN 21Q11-21 IN SQUAMOUS NON-SMALL CELL LUNG CARCINOMA AND CLONING OF A NOVEL UBIQUITIN-SPECIFIC PROTEASE GENE FROM THE DELETED SEGMENT" GENES, CHROMOSOMES & CANCER, XX, XX, vol. 27, no. 2, February 2000 (2000-02), pages 153-161, XP000943091 the whole document ---	1-7, 11-15
X	WO 99 46289 A (HUMAN GENOME SCIENCES INC ;NI JIAN (US); ROSEN CRAIG A (US); FERRI) 16 September 1999 (1999-09-16) SEQ ID NO 130 claims 1-23 ---	1-7, 9-18,25, 26
E	WO 00 78934 A (GROET JUERGEN ;NIZETIC DEAN (GB); UNIV LONDON PHARMACY (GB)) 28 December 2000 (2000-12-28) SEQ ID NOS 1,6 claims 1-35 ---	1-18
E	WO 00 79267 A (GROET JUERGEN ;NIZETIC DEAN (GB); UNIV LONDON PHARMACY (GB)) 28 December 2000 (2000-12-28) SEQ ID NO 1,6 claims 1-11 ---	1-7,9, 11-18
E	WO 01 21654 A (RIGEL PHARMACEUTICALS INC ;LUO YING (US); MANCEBO HELENA (US)) 29 March 2001 (2001-03-29) SEQ ID NO 1 claims 1-22; figure 2 ---	1-7, 9-14, 25-27
X	DATABASE EMBL [Online] Database entry HS1271038; Accession no AA479313, 23 June 1997 (1997-06-23) HILLIER L ET AL: "Homo sapiens cDNA clone" XP002176711 sequence ---	11,12
X	DATABASE EMBL [Online] Database entry HS1272487, Accession no AA482201, 24 June 1997 (1997-06-24) NATIONAL CANCER INSTITUTE, CANCER GENOME ANATOMY PROJECT (CGAP): "Homo sapiens cDNA clone similar to ubiquitin carboxyl-terminal hydrolase 2" XP002176712 sequence ---	11,12
		-/--

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/34811

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online] Database entry HSZZ12939, Accession no AA307805, 18 April 1997 (1997-04-18) ADAMS MD ET AL: "EST178661 Colon carcinoma (HCC) cell line Homo sapiens cDNA" XP002176713 sequence -----</p>	11,12

INTERNATIONAL SEARCH REPORT

Int'l. application No.
PCT/US 00/34811

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. Claims Nos.: **20, 21, 23, 24**
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

In part 1-28

Remark on Protest

The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20,21,23,24

Claim 20 refers to an agonist of a polypeptide of claim 1 and claim 23 refers to an antagonist of a polypeptide of claim 1 without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT).

No search can be carried out for such purely speculative claims whose wording is in fact, a mere recitation of the result to be achieved. The above comments also applies for claims 21 and 24.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/34811

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9946289	A	16-09-1999	AU	3006799 A	27-09-1999
			EP	1062236 A1	27-12-2000
			WO	9946289 A1	16-09-1999
			EP	1044210 A1	18-10-2000
			WO	9931116 A1	24-06-1999
			US	2001024813 A1	27-09-2001
<hr/>					
WO 0078934	A	28-12-2000	AU	5550800 A	09-01-2001
			AU	5552400 A	09-01-2001
			WO	0078934 A2	28-12-2000
			WO	0079267 A2	28-12-2000
<hr/>					
WO 0079267	A	28-12-2000	AU	5550800 A	09-01-2001
			AU	5552400 A	09-01-2001
			WO	0078934 A2	28-12-2000
			WO	0079267 A2	28-12-2000
<hr/>					
WO 0121654	A	29-03-2001	AU	7832800 A	24-04-2001
			WO	0121654 A2	29-03-2001
<hr/>					